



FUNDACIÓN UNIVERSITARIA
LOS LIBERTADORES

**DESARROLLO DE UN MODELO DE PRONÓSTICO EN LA METODOLOGÍA DE
BOX-JENKINS PARA LA DEMANDA DE ESTUDIANTES DE PREGRADO DE
LA FUNDACIÓN UNIVERSITARIA TECNOLÓGICO COMFENALCO DE LA
CIUDAD DE CARTAGENA DE INDIAS DESDE 2002 A 2018**

JUAN EMILIO TOBÓN GUZMÁN

**FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
DEPARTAMENTO DE CIENCIAS BÁSICAS ESPECIALISTA EN ESTADÍSTICA
APLICADA**

BOGOTÁ, D.C., 2019



FUNDACIÓN UNIVERSITARIA
LOS LIBERTADORES

**DESARROLLO DE UN MODELO DE PRONÓSTICO EN LA METODOLOGÍA DE
BOX-JENKINS PARA LA DEMANDA DE ESTUDIANTES DE PREGRADO DE
LA FUNDACIÓN UNIVERSITARIA TECNOLÓGICO COMFENALCO DE LA
CIUDAD DE CARTAGENA DE INDIAS DESDE 2002 A 2018**

JUAN EMILIO TOBÓN GUZMÁN

Trabajo de grado para optar al título de Especialistas en Estadística Aplicada

Asesor

Sebastien Lozano Forero

**FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
DEPARTAMENTO DE CIENCIAS BÁSICAS ESPECIALISTA EN ESTADÍSTICA
APLICADA**

BOGOTÁ, D.C., 2019



FUNDACIÓN UNIVERSITARIA
LOS LIBERTADORES

Nota de Aceptación

Presidente del Jurado

Jurado

Jurado

Bogotá, D.C., Junio de 2019



Índice de Contenido

1. RESUMEN.....	7
2. INTRODUCCIÓN.....	8
3. DESCRIPCION Y PLANTEAMIENTO DEL PROBLEMA.....	9
3.1 Descripción del problema.....	9
3.2 Objetivo general	13
3.3 Objetivos específicos	13
3.4 Justificación.....	13
4. MARCOS REFERENCIALES	16
4.1 MARCO TEORICO.....	16
4.1.1 Métodos de regresión lineal y múltiple.....	16
4.1.2 Métodos de suavizamiento exponencial	17
4.1.3 Métodos de proyección ARIMA	18
4.1.4 Métodos de transformación de datos.....	20
4.1.5 Pruebas de estacionariedad de series de tiempo	22
4.1.6 Modelo ARIMAX	23
4.2 Antecedentes	25
4.3 Marco metodológico	29
4.3.1 Método.....	29
4.3.2 Enfoque	31
4.3.3 Diseño metodológico	31
5. ANALISIS DE RESULTADOS	33
5.1 Identificación	33
5.2 Estimación del modelo	37
5.3 Verificación del diagnostico.....	38
5.4 Pronostico	39
6. CONCLUSIONES	41
7. REFERENCIAS	42



Índice de Gráficas

Gráfica 1. Demanda de estudiantes FUTCO 2012-2017.....	10
Gráfica 2. Metodología Box Jenkins.....	30
Gráfica 3 Diseño Metodológico	32
Gráfica 4 Serie de tiempo Matriculas 2001 1-2018 2 FUTCO	33
Gráfica 5 Distribucion de la serie de tiempo	34
Gráfica 6 Diagrama ACF y PACF serie de tiempo	35
Gráfica 7 Análisis serie de tiempo diferenciada.....	36
Gráfica 8 Análisis ACF y PACF de residuos.....	38
Gráfica 9 Datos originales Vs Datos ajustados	39



Índice de Tablas

Tabla 1 Prueba Jarque Bera	34
Tabla 2 Prueba de estacionariedad serie de tiempo	35
Tabla 3 Comparación y ajuste de modelos	37
Tabla 4 Prueba Jarque Bera residuos.....	38
Tabla 5 Prueba Ljung Box	39
Tabla 6 Pronostico 2019 1- 2021 2 serie de tiempo	40



1. RESUMEN

En este documento se presenta un acercamiento del modelo ARIMA y la metodología Box Jenkins para estimar un modelo de predicción de las matriculas académicas de una institución universitaria en la ciudad de Cartagena en su primer curso de pregrado. Se parte desde una contextualización del problema donde a nivel nacional y regional del departamento de Bolívar se tiene una tendencia a la baja de las matriculas en los últimos 3 años. Frente a este panorama se tiene como objetivo definir un modelo de pronóstico de la Fundación Universitaria Tecnológico Comfenalco para sus matrículas académicas de pregrado en su primer curso. El modelo de predicción se compone por una serie de 36 datos que parten desde el semestre 2001 1 hasta 2018 2. Una vez realizada la predicción del modelo se tiene un comportamiento de la predicción con datos estables para la demanda sin mostrar una disminución notable de la demanda en los próximos 3 años.

Palabras clave: modelo ARIMA, Box Jenkins, Demanda universitaria



2. INTRODUCCIÓN

El estudio relacionado con la aplicación de la metodología Box Jenkins para estimar la demanda estudiantil en una institución universitaria de Cartagena obedece a una necesidad que trasciende desde un fenómeno social que ha ocurrido en los últimos años, donde ha ocurrido una notable disminución de la demanda estudiantil, la cual ha escalado a un nivel regional como el departamento de Bolívar, su capital Cartagena de Indias y a las instituciones privadas, las cuales tienen una gran dependencia del ingreso generado por estas matriculas para su sostenibilidad.

La información recolectada para desarrollar este modelo corresponde al histórico contemplado entre los periodos de 2001 en su primer semestre hasta el periodo 2018 en su segundo semestre, contando con 36 registros que le dan una suficiente información para dar robustez al modelo que se desea generar. Esta información surge desde lo recolectado por información interna de la Fundación Universitaria Tecnológico Comfenalco, la cual es el objeto de estudio en el presente trabajo y de información tomada desde el SNIES (Sistema Nacional de Información de Educación Superior)

Seguido a esto el documento fundamenta teóricamente la metodología Box Jenkins y describe el marco metodológico para su aplicación. Igualmente se describe antecedentes que marcan la aplicación de esta herramienta. En una última etapa del documento se plantean los resultados del análisis caracterizando la serie de tiempo que es objeto de estudio, desde características de estacionariedad de esta para determinar el tipo de modelo ARIMA a construir.

Finalmente se realiza una estimación del modelo, así como su validación mediante pruebas de normalidad y estacionariedad de los residuos del modelo; así como la determinación de un pronóstico a 3 años para definir una estimación del alcance que tendrá la demanda a un largo plazo.



3. DESCRIPCIÓN Y PLANTEAMIENTO DEL PROBLEMA

3.1 Descripción del problema

En el contexto nacional se está registrando en los últimos 3 años un decaimiento de la población estudiantil en las universidades a un nivel de estudio de pregrado. Según cifras del SNIES, Sistema nacional de información de la educación superior, muestra que a nivel nacional hubo una caída de matrícula entre 2016 y 2017 de un 4,25%, pasando de tener 952.988 a 912.468 estudiantes. Esta disminución ha afectado tanto a IES (Instituciones de educación superior) públicas como privadas, las cuales han tenido una disminución del 3,61% y del 11,35% respectivamente.

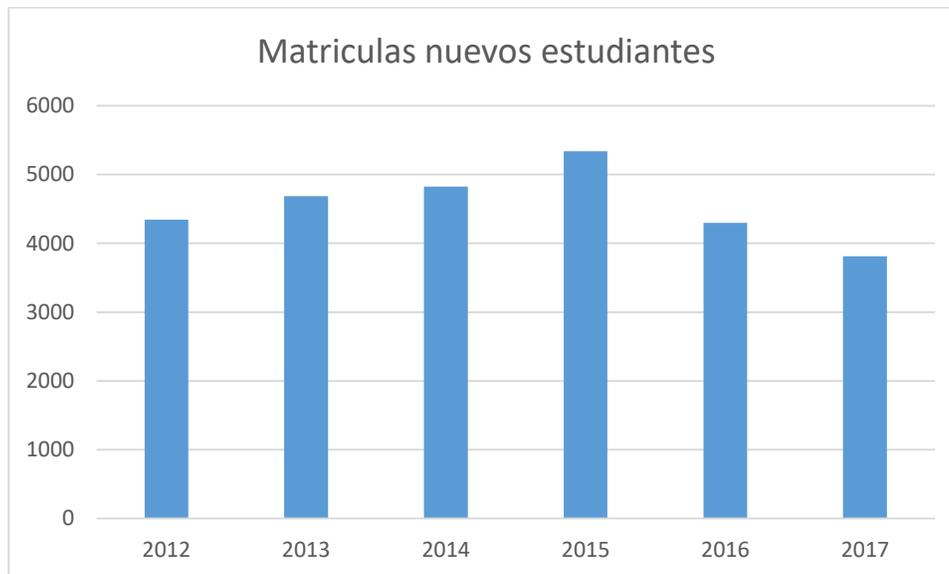
Esta última cifra se hace más alarmante en el sentido de que las instituciones privadas dependen en gran parte, directamente de los ingresos de matrículas para poder financiar sus operaciones y servicios de educación, investigación y proyección social, las cuales son las funciones sustantivas de las IES.

En el contexto de la ciudad de Cartagena, según datos del SNIES, la disminución de la población se reflejó pasando de una población de 20.851 estudiantes de pregrado en las IES en el año 2016 a 19.121 en el año 2017, teniendo una disminución del 8,29%, casi duplicando la media nacional durante la comparación de estos dos periodos. Para el sector de las IES públicas u oficiales se determinó una disminución de la población en un 0,74%, mientras que para las IES privadas significó un 11,57 en la disminución de su población en tan solo este cambio de periodos, el cual está un poco por encima de la media nacional.

Es en este último campo en el que se encuentra la Fundación Universitaria Tecnológico Comfenalco (FUTCO), entre 2016 y 2017 ha tenido una disminución de su población estudiantil del 11,32%, dicha cifra se es muy similar a las condiciones nacionales de las IES privadas y solo un poco por debajo de las condiciones de la



ciudad de Cartagena de Indias. A continuación se presenta un consolidado del comportamiento actual de la serie de tiempo a partir del año 2012.



Gráfica 1. Demanda de estudiantes FUTCO 2012-2017

Actualmente no se cuenta con cifras oficiales del SNIES del año 2018 sin embargo se sigue la tendencia que se marca desde el año 2015, donde se evidencia un descenso en la población de nuevos matriculados y por ende de la población estudiantil total en la FUTCO.

Las causas que se asocian con este comportamiento de la demanda estudiantil se representan en que:

El crecimiento económico de la ciudad se ha frenado desde la finalización del proyecto de la ampliación de la Refinería de Cartagena. El gran motor de la economía de la ciudad durante principios y mediados de la presente década fue la ampliación de la Refinería de Cartagena, la cual trajo una alta inversión y motivación por la población estudiantil que egresaba de niveles secundarios en estudiar carreras a nivel tecnológico e industrial que ofrece el FUTCO. Sin embargo con el cierre del proyecto y la notable reducción del personal en esas operaciones, hubo una severa reducción del movimiento económico en la ciudad que se vio reflejado en las pretensiones de los nuevos estudiantes de estudiar programas tecnológicos e industriales.



Crecimiento de la competencia en programas técnicos y tecnológicos en la ciudad. La estrategia de masificar e incentivar los servicios del SENA a nivel nacional, ha generado una gran atracción de los nuevos estudiantes de carreras tecnológicas y de pregrado hacia esa institución, lo que impacta directamente a la fluctuación de la demanda de la FUTCO

Preferencias demográficas a no estudiar. Las nuevas tendencias de la población de egresados de nivel secundario manejan un perfil o comportamiento a llevar a cabo una carrera laboral sin manejar una formación previa de un pregrado, gracias al acceso tecnológico y a nuevos formatos de formación mediante cursos rápidos y de formación específica.

Envejecimiento de la población. Según cifras del último censo del DANE en 2018, a nivel nacional se ha reflejado una disminución de la población estudiantil apta para estudios universitarios, pasando en 1973 de un 11,8% de la población a un 8,7% en 2018. Esto se debe a la preferencia de las familias de tener menos hijos en su núcleo familiar en las últimas generaciones.

Tendencias económicas de inflación e IPC. Según datos del Banco De La Republica en el año 2018 Cartagena fue la cuarta ciudad que tuvo una mayor inflación, teniendo una variación del 3,45% con respecto al año anterior, así también hubo un incremento de 0,28 puntos del IPC durante el mismo periodo, dando un total de 3,45%. Estas cifras indican una disminución del poder adquisitivo de la población de esta ciudad, que notablemente puede influir en las posibilidades de matrículas de nuevos estudiantes e incluso de antiguos, afectando especialmente a universidades del sector privado que dependen especialmente de las matriculas como principal medio de financiación.

Niveles de pobreza han aumentado en la ciudad. Según cifras del DANE, entre el periodo del 2012 y 2018 el ingreso de las personas que se consideran en niveles de pobreza se contrajo en un 37% en promedio, lo que ha generado una mayor desigualdad y pocas posibilidades para que gran parte de la población pueda acceder a estudios universitarios, especialmente porque el segmento de mercado



en que se concentra la FUTCO se encuentra en estratos económicos 1, 2 y 3, los cuales tienen una alta probabilidad de sufrir esa contracción de sus ingresos.

Ante estas causas el efecto que se presenta en la FUTCO se refleja especialmente en la incertidumbre del rumbo que tendrá la población de nuevos matriculados a la institución, lo cual es un insumo necesario para los procesos de planeación institucional y definición de objetivos a un largo plazo. Si se sigue presentando este fenómeno la institución seguirá presentando o presentara los siguientes síntomas o efectos:

Incertidumbre en la planificación y programación de recursos a largo plazo. Dentro de los procesos de planificación de la universidad se requiere la integración y adquisición de recursos a largo plazo como personal docente, equipos para laboratorio y ampliación de infraestructura. Sin embargo no se puede realizar una apuesta de inversión de este tipo sin tener una certeza del comportamiento de la población estudiantil semestre a semestre.

Replanteamiento de las estrategias de mercadeo de la institución. En caso de no tener una certeza del rumbo que tendrá el comportamiento de la demanda a un largo plazo, no es posible tener una preparación previa de las acciones que se tendrán en cuenta para atraer al mercado efectivamente a la propuesta de valor de la FUTCO, por lo que se hace relevante dicha información para estructurar una mezcla de mercadeo acorde a la población objetivo actual o a nuevos segmentos de mercados.

Despidos y cambios en la modalidad de contratación del personal docente y administrativo. Una incorrecta planificación de los recursos puede provocar una contratación de personal en exceso que genere sobrecostos o baja productividad en las operaciones del servicio, generando despidos o disminución de la modalidad de contratación, tal como pasar de un contrato a tiempo indefinido a horas catedra.

Sobrecarga académica del personal docente. En el caso contrario, cuando no se tenga el personal suficiente por el crecimiento de la población que no se tuvo prevista, se puede tener en este caso una sobrecarga académica en las funciones



asignadas de docencia. Generando improductividad en el servicio y descuido de funciones administrativas por parte del docente.

Teniendo en cuenta estos elementos del contexto se plantea la siguiente pregunta problema:

¿Cómo se proyecta la demanda de los nuevos matriculados en la población estudiantil de la Fundación Universitaria Tecnológico Comfenalco a partir de una serie de tiempo entre los años 2012 y 2019?

3.2 Objetivo general

Desarrollar un modelo de pronóstico en la metodología de Box-Jenkins para la demanda de estudiantes de pregrado de la Fundación Universitaria Tecnológico Comfenalco de la ciudad de Cartagena de Indias.

3.3 Objetivos específicos

1. Identificar las variables que hacen parte del modelo de pronóstico de serie de tiempo
2. Formular un modelo de pronóstico ARIMA a partir de la serie de datos de demandas de estudiantes
3. Validar el modelo cuantitativo de pronóstico a través de la metodología de Box Jenkins

3.4 Justificación

El propósito de este proyecto busca desde el punto de vista del autor, brindar y aportar el desarrollo de herramientas de pronósticos de demanda en campos microeconómicos definidos, que amplíen no solo la formación profesional, sino también la aplicación de nuevas herramientas que complementen los procesos de decisión de las IES, ya que dicho estudio puede ser replicable en otros entornos



que manejan las mismas condiciones a un nivel nacional y no existe un referente formal de dicho estudio. La aplicación de un modelo de serie de tiempo para el pronóstico de la demanda de nuevos estudiantes matriculados en la institución FUTCO permitirá realizar proceso de planeación y programación de recursos de forma coordinada y pertinente a las condiciones del mercado, de esta manera se puede trabajar con una mayor productividad en la asignación de recursos de procesos administrativos y operativos dentro del servicio educativo

La aplicación metodológica de Box Jenkins ha tenido diferentes trabajos de aplicación donde se presente su respectiva validación, tal como lo demuestra MUNENAKA & CARMEN (2017)¹, donde se hace una revisión de los diferentes modelos de pronóstico y aplicados a una demanda comercial, se valida la pertinencia de un ARIMA para reducir el error de estimación de la serie de tiempo. Ante ello la pertinencia de una revisión literaria sobre las diferentes metodologías dentro del estudio actual para dar validación de la metodología Box Jenkins sobre otros modelos de pronósticos.

Teniendo en cuenta esto se toma como referencia la aplicación de estas herramientas de pronóstico y la metodología Box Jenkins para la estimación de la demanda fuera del entorno microeconómico de las IES, para eso se tiene en cuenta lo aportado por OLIVERA, DOURADO & MELLO (2016)², donde aplican la metodología Box Jenkins para identificar un modelo de pronóstico que se adecue al comportamiento de la demanda de una empresa en el sector plástico. Para llegar a ello tuvieron en cuenta la estacionalidad de la demanda, en un periodo de 12 meses, la cual se aplica de igual manera para la IES del caso de estudio de este proyecto, en este caso segregado en 2 semestres. El resultado del proyecto generó un modelo de pronóstico ARIMA que generó un mínimo de errores en el proceso de validación,

¹ Munenaka, E., & Carmen, V. (2017). Pronóstico de demanda utilizando la metodología de Box-Jenkins.

² de OLIVEIRA, E. S., de Almeida DOURADO, J. D., & MELLO, J. A. V. B. (2017). APLICAÇÃO DE MODELOS DE PREVISÃO DE DEMANDA EM UMA FÁBRICA DE EMBALAGENS PLÁSTICAS. *Revista da Universidade Vale do Rio Verde*, 15(2), 354-373.



el cual permitió dar en ese caso la información pertinente para poder generar un proceso de planeación acorde a la demanda.

PAREDES (2018)³ desarrolla igualmente la aplicación de la metodología Box Jenkins, en este caso para trabajar sobre una demanda afectada por variables que se encuentran fuera del entorno, tal como los turistas extranjeros. La aplicación de esta metodología ofrece un modelo de pronóstico SARIMA, que indica un ascenso de esta demanda hacia un futuro lejano, lo que le permite a las autoridades realizar un proceso de planeación y ampliación de la cobertura de servicios turísticos.

Igualmente el estudio se justifica al emplear la metodología Box Jenkins en procesos de pronóstico para la demanda de un servicio energético con el fin de realizar una planificación operacional de los procesos de producción confiable y eficiente a partir del estudio de ESPINOZA, VICENTE & SEMPÉRTEGUI (2017)⁴. Con esta breve recopilación de estudios, se verifica la pertinencia de la aplicación de la metodología Box Jenkins como pertinente y justificable para el caso de estudio planteado en la presente propuesta de trabajo de investigación.

³ Mamani, R. P. P. (2018). Proyección de la demanda de Turismo Internacional en Puno: Un enfoque SARIMA. *Semestre Económico*, 5(2).

⁴ Espinoza, M., Vicente, S., & Sempértegui Cañizares, E. R. (2017). *Proyección a corto plazo de la demanda eléctrica de los alimentadores de distribución de la Empresa Eléctrica Regional Centro Sur CA* (Master's thesis, Universidad del Azuay).



4. MARCOS REFERENCIALES

4.1 MARCO TEORICO

4.1.1 Métodos de regresión lineal y múltiple

El análisis de regresión lineal y análisis de regresión lineal múltiple son procedimientos cuantitativos que permiten establecer la relación que se produce entre una variable dependiente y una o un conjunto de variables independientes respectivamente. La regresión lineal se diferencia de la múltiple en que esta última puede contar con una mayor aproximación a situaciones reales y contextualizadas al comportamiento de diferentes variables que ofrecen los entornos a analizar. Este modelamiento matemático y estadístico busca entonces explicar la complejidad de hechos y fenómenos sociales a través de variables que intervienen en su ejecución (Rodríguez & Mora, 2001)⁵

En la regresión lineal múltiple se utiliza entonces más de una variable explicativa, por lo que ofrece una ventaja a un analista para plantear con mayor información la proyección de una variable independiente a partir de una dependiente. Sin embargo dentro del conjunto de variables independientes es papel del analista y de los procedimientos cuantitativos de la regresión lineal múltiple entender que tanto influye cada una de estas sobre la proyección deseada.

En consecuencia se considera entonces la relación de las variables independientes con la variable dependiente mediante un modelamiento matemático que implica la presencia de coeficientes determinados de forma que la suma de cuadrados entre los valores observados y los pronosticados sea mínima, es decir, que se va a minimizar la varianza residual (Abuin, 2007)⁶

⁵ Rodríguez-Jaume, M. J., & Mora Catalá, R. (2001). Análisis de regresión múltiple. *Técnicas de Investigación Social II*.

⁶ Abuín, J. R. (2007). Regresión lineal múltiple. *IdEyGdM-Ld Estadística, Editor*, 32.



$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \dots + b_kx_k + u$$

Esta ecuación se conoce como hiperplano y se maneja en un orden igual al número de variables independientes.

Un modelo de regresión lineal múltiple debe manejar y cumplir con hipótesis de linealidad, homocedasticidad, independencia, normalidad y las variables explicativas se obtienen sin errores de medidas. El cumplimiento de cada una de estas consideraciones explica que el modelo de aproximación mediante mínimos cuadrados ofrece estimadores óptimos para el modelo matemático.

Es entonces papel del analista determinar y ajustar el modelo de regresión lineal múltiple que reduzca especialmente la varianza residual de cada una de las proyecciones del modelo, es decir, que la distancia entre la proyección y los datos reales en la validación debe ser la mínima posible. Lo que implica que la modelación matemática debe someterse a un proceso de ajuste de sus coeficientes que dan pesos a la representación de cada una de las variables independientes para dar cabida a una situación óptima del modelo matemático final que cumpla con las pruebas de hipótesis anteriormente planteados.

4.1.2 Métodos de suavizamiento exponencial

El método de suavizamiento exponencial simple se considera como la evolución del método de promedio móvil ponderado aplicado a una serie de tiempo representado por una variable. Este método se considera un mecanismo de autocorrección que ajusta los pronósticos en dirección a las desviaciones del pasado a través de un coeficiente de ajuste y suavización.

De esta manera el pronóstico trabajado mediante este método requiere entonces de tres elementos de entrada: el pronóstico del último periodo, el último dato de la serie de tiempo y el coeficiente de suavización sobre este último.

El método de suavizamiento exponencial se caracteriza busca entonces ponderar con mayor peso a las observaciones más recientes y con menor a las antiguas,



basado en el planteamiento de que los valores más recientes tienen una mayor correlación con el pronóstico. Dentro de estos métodos existen varios tipos: únicos, doble lineal, lineal y estacional y amortiguado de tendencia. De esta manera la ecuación que define este pronóstico simple viene asignado por:

$$F_{t+1} = \alpha X_t + (1-\alpha)F_t$$

El pronóstico de suavización exponencial simple es óptimo para patrones de demanda aleatorios tal como funcionan una serie de tiempo bajo una caminata aleatoria o nivelados bajo consideración de un ruido blanco donde se pretende eliminar el impacto de los elementos irregulares históricos mediante un enfoque en períodos de demanda reciente, este posee una ventaja sobre el modelo de promedio móvil ponderado ya que no requiere de una gran cantidad de períodos y de ponderaciones para lograr óptimos resultados.

Esta especial utilidad del método de suavizamiento exponencial permite que sean involucrados en proyectos de corto plazo, donde no se cuenta con un historial de datos que sean requeridos para la identificación de un pronóstico. Los pronósticos a corto plazo tienden a ser más precisos que los de largo plazo. Los factores que influyen en la demanda cambian todos los días, por tanto, en la medida que el horizonte de tiempo se alarga, es más probable que la precisión del pronóstico disminuya (Render & Heizer, 2007)⁷

A partir del método de suavizamiento exponencial se puede hacer una extrapolación de la aplicación a un contexto de las ventas y proyección de la demanda, basado en un histórico amplio de datos, llegando a la metodología Box-Jenkins.

4.1.3 Métodos de proyección ARIMA

A principios de los años 70 Box y Jenkins desarrollaron una metodología que permiten identificar, estimar y diagnosticar modelos de series temporales, especialmente desde un perfil econométrico por parte del analista.

⁷ Render, B., & Heizer, J. (2007). Administración de la Producción. México: Hugo Rivera Oliver.



El procedimiento de esta metodología inicia con un análisis a partir del mismo pasado que protagoniza a la variable que se desea proyectar. Esto implica una ventaja ya que no se requiere de una variable exógena para poder realizar una proyección hacia el futuro, sin embargo esta única dependencia puede ser vista como un inconveniente, y es que no se atiende ninguna relación con otra variable que se encuentre dentro del entorno de la variable que se está analizando, bajo la consideración de un modelo ARIMA, incluso con un componente estacional.

Este modelo ARIMA se considera entonces como univariante, por sus siglas como un modelo autorregresivo integrado de medias móviles. La estimación de este modelo se integra bajo la metodología Box Jenkins, el cual toma en cuenta el comportamiento de la serie de tiempo como un proceso estocástico y con estacionariedad, representado por Y_k datos a través de tiempo. Siendo este último el elemento contemporáneo y desde donde se parte para encontrar una correlación con el pasado de esta misma variable pero con un futuro muy incierto, es decir que esta proyección se asociara entonces con una distribución de probabilidad a su posible comportamiento dentro del modelo ARIMA, de ahí su comportamiento estocástico.

Mientras que su estacionariedad se ve representada por el comportamiento de ruido blanco en una serie de tiempo de una variable aleatoria, un proceso estocástico, cuya media poblacional es de cero, tiene una varianza constante y covarianza de cero. Esto generalmente se conoce como estacionariedad fuerte o estricta (Arce & Mahía, 2003)⁸

Mientras que una estacionariedad débil se interpreta cuando tanto la media poblacional así como su varianza y covarianza no dependen del tiempo, es decir serán constantes, finitas y dependerán del lapso de la serie de tiempo, respectivamente. Para definir el modelamiento de una serie de tiempo a través de la metodología Box Jenkins se busca asegurar que esta tenga un comportamiento de estacionariedad, ya que entonces sus variables pueden estar relacionadas

⁸ De Arce, R., & Mahía, R. (2003). Modelos Arima. *Programa CITUS: Técnicas de Variables Financieras*.



linealmente entre si, pero de forma que la relación entre dos variables sólo depende de la distancia temporal k transcurrida entre ellas (De Arce & Mahía, 2003)⁹.

El componente autorregresivo de un modelo ARIMA explica la dependencia de la variable que protagoniza la serie de tiempo en un momento k contemporáneo con t observaciones hacia el pasado, considerando que dicha variable sigue un proceso estocástico y estacionario, la relación de la observación contemporánea con su pasado sería lineal mas un componente de error. Siendo así el componente autorregresivo AR de orden k se estructura matemáticamente como

$$Y_t = \phi_0 + \phi_1 Y_{t-1} \dots + \phi_k Y_{t-k} \dots + \varepsilon$$

Mientras que el componente de medias móviles MA se compone por el valor de una determinada variable en un período t en función de un término independiente y una sucesión de errores correspondientes a períodos precedentes, ponderados convenientemente. Esto significa que una variable aleatoria que representa a la serie de tiempo tendrá una proyección explicada en función a los errores cometidos en periodos precedentes.

4.1.4 Métodos de transformación de datos

Ante un comportamiento de estacionariedad débil de una serie de tiempo se hace utilizar su comportamiento hacia un ruido blanco o a una serie de tiempo con estacionariedad estricta y fuerte. Esto es importante para la aplicación de modelos con componente autorregresivo y de medias móviles (ARMA). La intención de lograr la estacionariedad radica en obtener una media y varianza constante a través de la serie de tiempo, para ello se han desarrollado procedimientos matemáticos que realizan transformaciones de los datos originales de la serie de tiempo. Estos son:

- Transformaciones de estabilización de la varianza: para este tipo de transformación se suele aplicar generalmente logaritmos naturales y de raíz

⁹ Et al



cuadrada, aplicados directamente a los valores de la serie. Esto se hace en series de tiempo en la que la varianza cambia a lo largo del tiempo.

- Transformación de estabilización de nivel de la media: en este caso se entiende que los datos de la serie de tiempo están estrechamente correlacionado a través del tiempo, lo que no se asemeja a un ruido blanco o comportamiento aleatorio de la variable a analizar. Para corregir dicho comportamiento se aplica el método de diferenciación, el cual consiste en realizar una diferencia algebraica entre cada uno de los valores existentes de la serie con su respectivamente anterior dato. En caso de que se determine un comportamiento estacional en la serie de tiempo, se aplica una diferenciación estacional

Si se usa la diferenciación simple o estacional de forma simultánea con la transformación logarítmica o de raíz cuadrada, siempre se aplicará primero la transformación de estabilización de la varianza. Si se usan la diferenciación simple y estacional, los valores de la serie resultante son iguales independientemente de si se aplica primero una diferenciación u otra.

- Falta de normalidad y heterocedasticidad: Uno de los procedimientos más usados para resolver estos problemas, es el transforma la serie de tiempo a través de la utilización de la familia de transformaciones BOX-COX. Para ello Se desea transformar la variable Y que representa a la serie de tiempo, cuyos valores se suponen positivos, en caso contrario se suma una cantidad fija M tal que $Y + M > 0$. La transformación de Box-Cox depende de un parámetro por determinar y viene dada por una lambda que permita darle la mayor verosimilitud a la serie de tiempo. Este lambda debe cumplir con las siguientes condiciones

$$Z(y) = \begin{cases} \frac{y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(y) & \lambda = 0 \end{cases} \quad y > 0$$



4.1.5 Pruebas de estacionariedad de series de tiempo

Las pruebas de estacionariedad consiste en un conjunto de procesos matemáticos y de inferencia estadística centrado especialmente en la diferenciabilidad de una serie de tiempo y la existencia o no de una raíz unitaria. Es importante reconocer que la existencia de una raíz unitaria en una serie de tiempo implica que esta no sea estacionaria y los estimadores del modelo matemático planteado para esta no tienen entonces una distribución normal y con un comportamiento a la deriva, por tanto si la serie tiene una raíz unitaria, ésta presenta un patrón sistemático que es impredecible¹⁰.

Como se ha mencionado existen procedimientos de diferenciación de la serie de tiempo para reducir la no estacionariedad de la serie de tiempo, así como también transformación de datos con el fin de eliminar la existencia de raíces unitarias. Las pruebas de estacionariedad o de raíces unitarias pueden ser usadas para determinar si una serie con tendencia primero deber ser diferenciada o hacer una regresión tendencial, esto es, si hay estacionariedad tendencial o diferencial. Los procedimientos utilizados para determinar el orden de integrabilidad de una variable son de dos tipos: los empleados en la metodología Box-Jenkins y los procedimientos basados en contrastes. Los primeros, en los que se basaba el análisis tradicional de series temporales, consisten en el examen gráfico de la serie y de los correlogramas (funciones de autocorrelación simple y parcial de la serie en cuestión). Para este primer método se esperaría que en el caso de series estacionarias, la función de autocorrelación declina rápidamente, al contrario de lo que sucede en presencia de una raíz unitaria que lo hace de forma irregular y prolongada. Este procedimiento, a pesar de su cómoda y fácil implementación no presenta la formalidad requerida y sus resultados pueden en muchos casos ser interpretados discrecionalmente. Por ello, además de examinar la varianza de la serie con distintos órdenes de diferenciación, se han ido planteando diversos

¹⁰ Castrillo-Rojas, D., & Rodríguez-Vargas, A. (2009). *Procedimiento para la aplicación de pruebas de raíz unitaria*. Banco Central de Costa Rica.



contrastes de raíces unitarias, algunos de los cuales (los más utilizados) se explicaran más adelante.

Las primeras pruebas de raíces unitarias sobre una serie de tiempo fueron desarrolladas por Dickey y Fuller (1979 y 1982) conocida como la prueba ADF (Dickey-Fuller aumentado), la cual elimina la autocorrelación e indica si una serie es estacionaria o no; más adelante se marca un desarrollo por Phillips Perron (1988) con su respectiva prueba PP, la cual corrige la autocorrelación y heterocedasticidad en los errores residuales. En estos test se propone como hipótesis nula la existencia de raíces unitarias, por lo tanto la hipótesis alternativa corresponde a la estacionariedad de la serie de tiempo¹¹.

Entonces con la prueba ADF se prueba la hipótesis nula de un proceso con raíz unitaria y constante versus la hipótesis alterna de que el proceso generador de los datos es un proceso estacionario alrededor de una tendencia, para ello se inicia probando la hipótesis nula, empleando un estadístico t convencional, pero, comparándolo con los valores críticos especiales. Un problema práctico que se presenta al momento de realizar esta prueba es determinar el número de rezagos p de manera previa. Para la prueba de PP no es necesario determinar un número de rezagos para el componente autorregresivo, pero, si es necesario determinar el número de rezagos de truncamiento.

4.1.6 Modelo ARIMAX

Un modelo ARIMAX con parámetros (p, d, q, r) Se considera un modelo general de modelos autorregresivos multivariados. La representación de este tipo de modelo es capaz de representar procesos lineales estacionarios y procesos lineales no estacionarios homogéneos con entradas exógenas. Se utiliza para series cuyos valores reales dependen de sus valores pasados, y los valores pasados con una

¹¹ Asteriou, D. (2002). Notas sobre Análisis de Series de Tiempo: Estacionariedad, Integración y Cointegración. *Obtenido de Notes on Time Series Analysis: \Stationarity, Integration and Cointegration: <http://www.personal.rdg.ac.uk/~less00da/lecture3.htm>*.



variable de entrada exógena más un ruido aleatorio en el instante actual (ruido blanco). Esta representación es presentado por la siguiente estructura polinómica:

$$\sum_{j=0}^p d_j \Delta^d y_{k-j} = b_0 x_k + \sum_{j=1}^r b_j \Delta^d x_{k-j} + \sum_{j=0}^q c_j a_{k-j}$$

Del modelo lineal ARIMAX se derivan varios modelos lineales, cuando las series son estacionarias, se puede utilizar una representación más precisa para el modelado llamado ARMAX, es decir que no cuenta con un componente diferenciador. En el caso de que se considere la estacionariedad en el serie de datos, se puede obtener un modelo de autorregresión de entrada exógena (ARX) o un modelo de media móvil con entrada exógena (MAX). Sin embargo, cuando no hay Estacionalidad en la serie, se utilizan los modelos integrados, como el modelo autorregresivo integrado de media móvil con entradas exógenas (ARIMAX), autorregresivo integrado con entradas exógenas (ARIX) y el modelo integrado de media móvil con entradas exógeno (IMAX).

La determinación y selección de una de las anteriores estructuras hace parte de la estructura del modelo final. Dentro de las diferentes opciones de identificación se emplea la metodología de Box y Jenkins (1970). Tal método consiste en utilizando la función de autocorrelación (ACF) para determinar el grado del polinomio promedio Función de autocorrelación parcial (PACF) para determinar el grado del polinomio autorregresivo En este sentido, las siguientes secciones explorarán la teoría de estas funciones y la función de Correlación Cruzada (FCC) utilizada en la validación del modelo.

Existen cuatro supuestos estadísticos que se debe examinar para asegurar que los resultados del modelo ARIMAX sean válidos, que se explican a continuación:



Supuesto 1: Se requiere que la serie de tiempo sea estacionaria, es decir, que media y varianza no varíen en el tiempo.

Supuesto 2: Los residuos no deben poseer un nivel significativo de autocorrelación. Lo evaluaremos mediante el test de Breush-Godfrey. Si existe presencia de este problema, se lo reduce agregando una combinación apropiada de uno o más términos AR y/o MA

Supuesto 3: El coeficiente estimado de una variable exógena debe ser significativo en su valor-p. El nivel significativo del valor-p de los coeficientes asume que los residuos de la regresión sean ruido blanco.

Supuesto 4: Una variable exógena puede mostrar una relación significativa con la variable dependiente, pero esto no se debe cumplir en viceversa. El nivel de causalidad lo mediremos usando el test de Granger . Si se observa una causalidad inversa, se debe retirar la variable independiente.

4.2 Antecedentes

Nyoni Thabani (2018) en su trabajo orientado a Aplicar la metodología Box Jenkins enfocado a la predicción de la entrada de inversión extranjera directa en Zimbabwe, realiza una proyección de la entrada de inversión extranjera en Zimbabwe, partiendo desde datos históricos de 1980 hasta 2017, aplicando la metodología Box Jenkins para desarrollar un modelo que permitiera visualizar la predicción en un horizonte de 20 años. Para ello aplica prueba ADF de estacionariedad en la serie de tiempo que le permite verificar que el modelo que reduce los residuos entre el modelo y el comportamiento real es un ARIMA(1,1,1), igualmente se aplica un diagnostico en las raíces inversas de los polinomios AR y MA que confirman la estabilidad del modelo. Las conclusiones del estudio y del modelo desarrollado permite al investigador generar una proyección de la serie de tiempo hacia 20 años, mostrando una tendencia no favorable para la inversión extranjero, por lo que se recomienda a partir del estudio idear políticas favorables para las inversiones extranjeras con el



fin de incrementar este indicador. Es importante rescatar de este antecedente la aplicación de pruebas ADF de existencias de raíces unitarias con el fin de validar la estacionariedad de la serie de tiempo, esto permite dentro de la metodología Box Jenkins entender la naturaleza y comportamiento de esta. Igualmente se encuentra la pertinencia de evaluar las raíces inversas de los coeficientes que conforman los polinomios AR y MA para validar el peso de cada uno de ellos dentro del modelo propuesto finalmente. Los pasos de la metodología Box Jenkins empleada por el autor tienen una menor complejidad a la tradicional, obviando una transformación de los datos para ajustar el modelo.

Oliveira, Dourado, y Mello (2017), en su trabajo a Seleccionar, aplicar y evaluar modelos de predicción de la demanda en una fábrica de embalajes plásticos a través de la metodología Box Jenkins, buscó con el estudio llevar a cabo la aplicación de la metodología Box Jenkins a la predicción de la demanda como insumo a los procesos de planeación y control de la producción en una fábrica de embalajes plásticos. Para ello se evaluaron diferentes modelos ARIMA, modelos con suavizamiento exponencial y modelos de medias móviles y se selecciona aquel donde se represente un menor error residual, a través de los indicadores MAPE, MAD y MSD. La evaluación del modelo del estudio permite seleccionar la metodología Box Jenkins con un modelo ARIMA (3,1,3) como aquel con menor error residual. Con la predicción propuesta de este modelo la empresa que se toma como caso de estudio puede inferir en sus procesos de planeación y control de la producción, lo que le permite establecer y alcanzar metas de ventas razonables, disminución de no conformidades y disminución de costos de operación. Este referente permite tener la metodología Box Jenkins como herramienta para la planificación de la producción. Incluso por encima de métodos de predicción como medias móviles, suavizamiento exponencial simple, suavizamiento exponencial doble, suavizamiento winters multiplicativa y aditiva. Se puede emplear para este comparativo, indicadores tales como MAPE(error porcentual absoluto medio), MAD(error absoluto medio) y MSD(error cuadrático medio).



Paredes y Paz (2018) en su estudio buscan Modelar y pronosticar la demanda turística de Puno utilizando series temporales mensuales para el periodo comprendido entre enero de 2003 y diciembre de 2015 empleando el modelo estacional de Box Jenkins, SARIMA. La pertinencia del estudio nace en referencia al crecimiento de la población turística dentro de la localidad objeto de estudio producto de estrategias de incentivo, sin embargo no se tiene una medida de control sobre el respectivo crecimiento, para ello se plantea un modelo de predicción mediante la metodología Box Jenkins con un componente estacional, SARIMA, debido al comportamiento de la demanda turística. El estudio detecta quiebres en la estacionalidad de la serie en 2 puntos de su cronología, octubre de 2010 y febrero de 2012.

Para determinar la estructura del modelo SARIMA se emplea un modelo multiplicativo mediante el método de media móvil ponderada exponencial doble DEWMA. Acorde a los resultados de la predicción se encuentra que el incremento de la llegada de los turistas se incrementara de manera moderada y las autoridades locales deben tomar medidas y estrategias para controlar el crecimiento de esta población de turistas. Este antecedente permite reconocer que un modelo SARIMA aplicado mediante la metodología Box Jenkins se puede integrar con un modelo de media móvil ponderada exponencial doble, con el fin de detectar cambios en la media del proceso.

Chong y Aguilar (2016) en su trabajo buscan presentar una proyección del consumo de electricidad para clientes residenciales del Ecuador para los meses que van desde el año 2013 hasta el año 2022; usando series de tiempo mediante la construcción de modelos ARIMA para el caso univariado y modelos ARIMAX para el caso multivariado. Los autores presentan la importancia de definir un modelo de predicción bajo la metodología Box Jenkins, llegando a un modelo ARIMA y a un modelo ARIMAX. Estos se emplearon respectivamente para series de tiempo univariadas y multivariadas. En base a esas predicciones se puede realizar una correcta planificación de los recursos requeridos para la producción energética de una nación. El estudio en su importancia y aportación al presente propone la



ampliación de la metodología Box Jenkins aplicadas para modelos ARIMA, y lo extiende a un modelo ARIMAX, explorando especialmente la existencia de variables exógenas significativas.

Silva (2016) en su trabajo busca analizar estadísticamente el comportamiento y evolución de los rendimientos obtenidos por área cultivada de arroz paddy en Colombia, durante un periodo de 35 años. El autor propone la aplicación de para lo cual se propone la aplicación de series de tiempo a los datos disponibles, específicamente un modelo ARIMAX (p,d,q) a través del cual se busca recoger la sensibilidad del rendimiento histórico y la incorporación de algunas variables exógenas como: el consumo de fertilizantes, el nivel de las reservas de capital en maquinaria agrícola, la temperatura media nacional y el indicador de precipitación nacional. El modelo obtenido es usado para describir el comportamiento de la serie y para proyectar el comportamiento esperado del rendimiento por hectárea del cultivo de arroz paddy de los próximos tres años. A partir de este estudio se tiene un antecedente de la validez y efectividad en un modelo Arimax sobre el ARIMA en la predicción de una variable, basándose en pruebas tales como Box Ljung, Jarque Bera, Shapiro Wilk y coeficientes BIC y AIC.

Romero (2017) en su estudio presentar un acercamiento a los modelos ARIMAX y lineal dinámico para estimar natalidad en Colombia en el periodo entre enero 2011 a diciembre 2017, El estudio parte de la necesidad que tiene el estado de contar con información poblacional veraz y confiable. Para el análisis exploratorio de los datos se realizan series de tiempo de los nacimientos en Colombia usando las variables edad de la madre y del padre, número de hijos y defunciones fetales; posteriormente, se analiza la serie nacimientos con las variables de apoyo, usando inicialmente todas las establecidas (edad madre, edad padre, hijos y defunciones fetales), quedando dos como variables exógenas significativas en el modelo ARIMAX. Luego del planteamiento de los modelos ARIMAX y el modelo lineal dinámico, como primera conclusión puedo señalar la fiabilidad de ambas estimaciones, los modelos se mostraron suficientes para explicar características de la serie de natalidad, en la cual pudieron estimar su tendencia y estacionalidad.

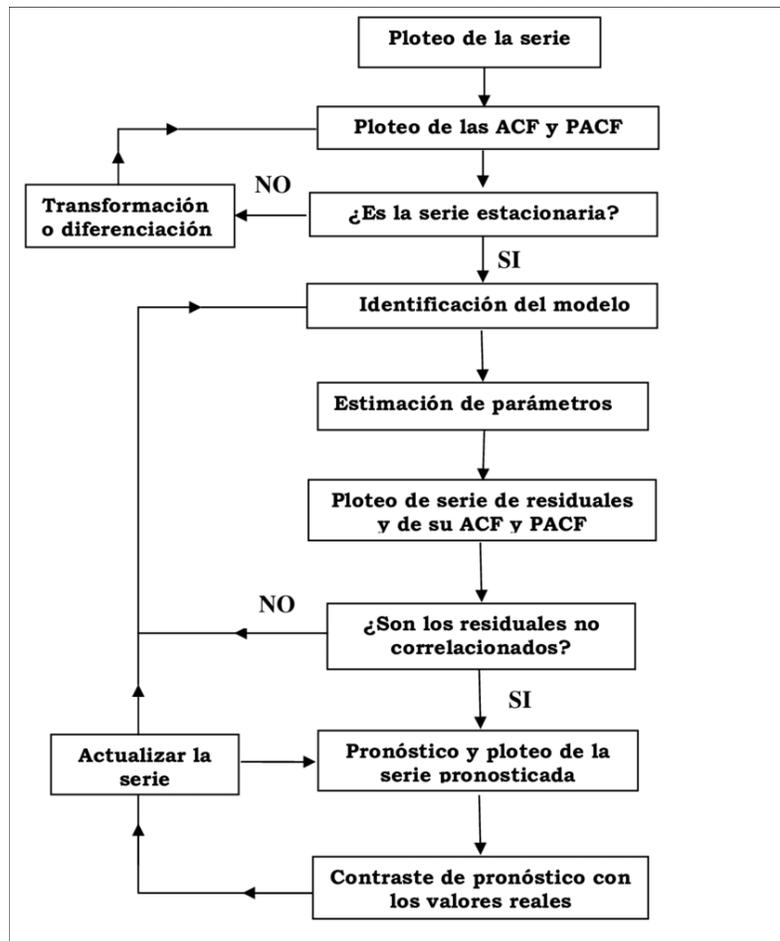


Molano (2016) busca en su trabajo analizar la serie que refleja el comportamiento del fraude mensual con tarjetas para una Entidad Financiera, así como también realizar un modelo ARIMAX(p,d,q) e identificar cual podría ser el fraude para el año 2016. El autor En busca de Analizar el mismo para evaluar mejores prácticas, demostrando la capacidad de reacción del banco y su posición frente a las otras entidades del sector. Este también aplica la validación de múltiples variables exógenas para llegar a una que se ajuste al modelo de predicción

4.3 Marco metodológico

4.3.1 Método

El método de solución de la problemática planteada será mediante la metodología Box Jenkins, la cual se ilustra en la siguiente grafica



Gráfica 2. Metodología Box Jenkins

Esta es una metodología iterativa que busca principalmente el ajuste de un modelo de pronóstico de una serie de tiempo con el fin de que este sea el más acertado. La metodología inicialmente se compone de tres etapas, la primera es la identificación y selección del modelo la cual se basa a partir del análisis de los diagramas de función de autocorrelación ACF y autocorrelación parcial PACF, asegurando de que el comportamiento de la variable sea estacionaria y se identifique también un componente de estacionalidad.

Una segunda fase consistirá en la estimación de los parámetros a través de la aplicación de algoritmos para la obtención de coeficientes del modelo lineal que de un mejor ajuste del modelo a las condiciones reales de la variable. Esto se puede hacer mediante estimación de máxima verosimilitud o por mínimos cuadrados no lineales. Y por último, una tercera fase que consiste en la validación y ajuste del



modelo a través de pruebas de independencia de los residuos, cuyas medias y varianzas deben ser constante en el tiempo, o especialmente manejar una distribución normal.

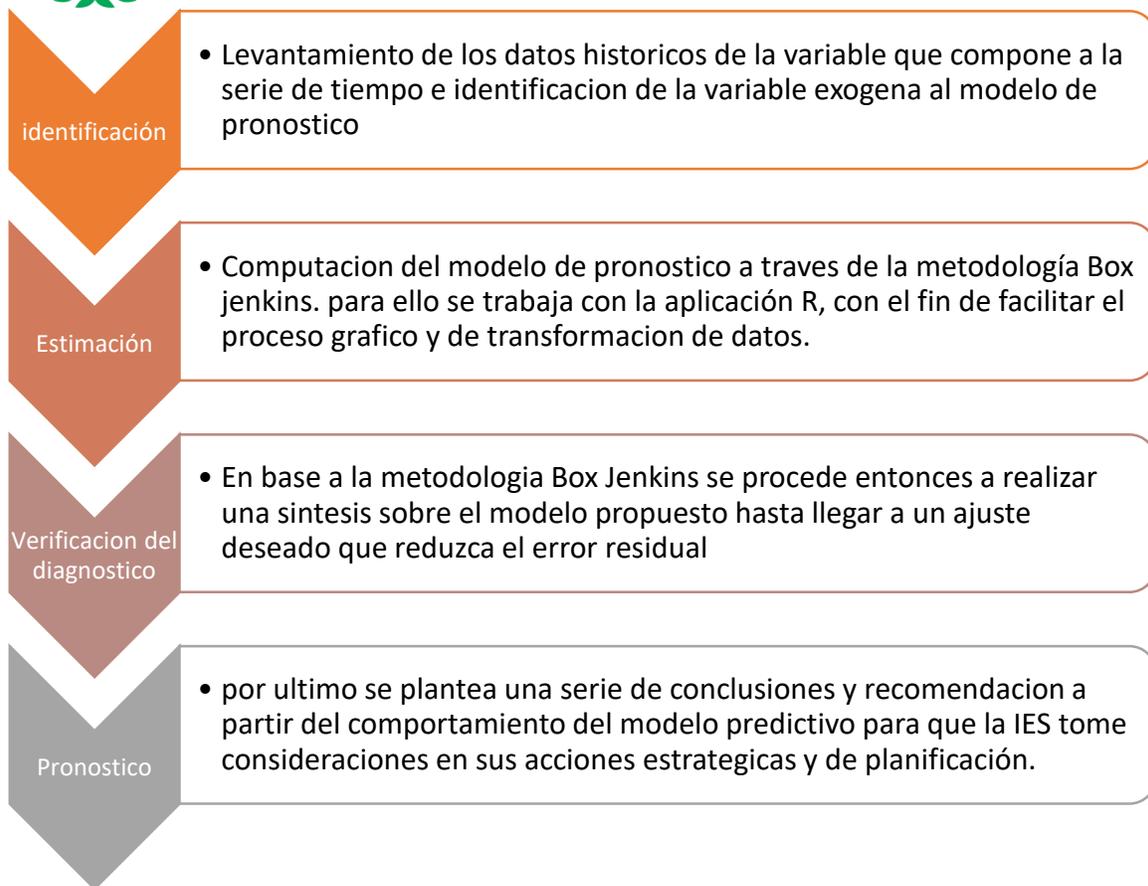
4.3.2 Enfoque

El enfoque de la investigación que se plantea es de tipo descriptiva y cuantitativo, la primera característica obedece a la intención de describir las características de un fenómeno presentado en un contexto limitado y definido por el investigador. Este enfoque se hace pertinente debido a que se tiene poca información del objeto de estudio a investigar y de un modelo de predicción que permita identificar su prospección en el tiempo.

El estudio guarda también un enfoque cuantitativo debido a que se trabaja con un proceso secuencial y riguroso partiendo desde una idea que ha sido delimitada a través de una problemática acompañado con un tratamiento objetivo de datos e información objetiva. Teniendo en cuenta que lo que se busca además es la medición de una variable a través de un proceso estocástico de una serie de tiempo representada por una variable y establecer a partir de ello una serie de conclusiones que guarden la objetividad del caso de estudio.

4.3.3 Diseño metodológico

Las actividades y etapas del estudio a realizar tienen en cuenta el siguiente proceso:



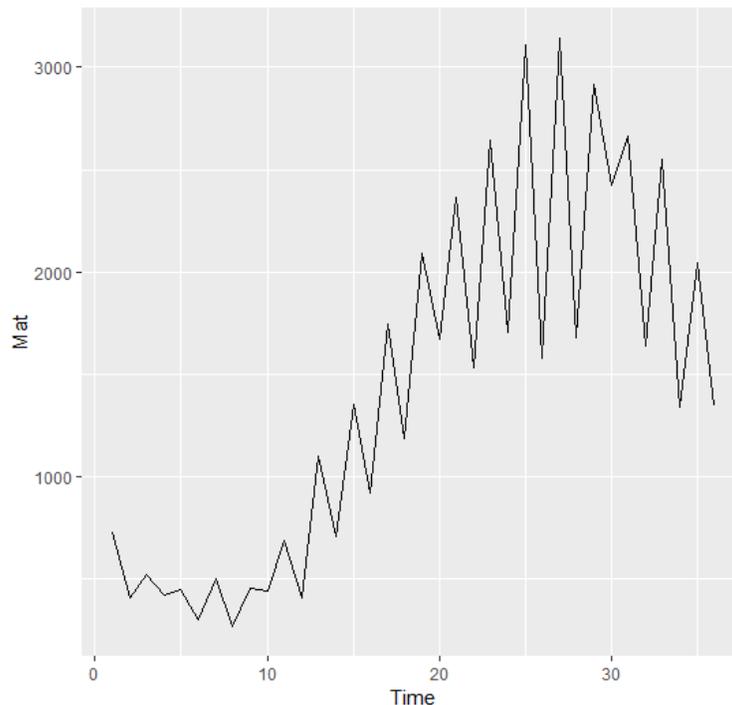
Gráfica 3 Diseño Metodológico



5. ANALISIS DE RESULTADOS

5.1 Identificación

La serie de tiempo de matrículas de estudiantes nuevos a los programas de pregrado de la Fundación Universitaria Tecnológico Comfenalco presenta el siguiente comportamiento.



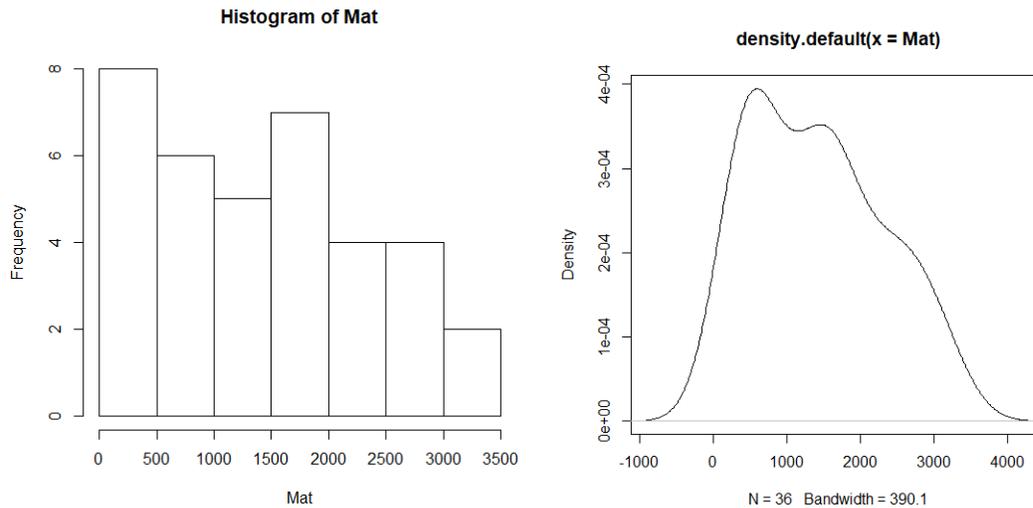
Gráfica 4 Serie de tiempo Matriculas 2001 1-2018 2 FUTCO

Se encuentra caracterizada por contener el histórico de estas matrículas por semestre desde el periodo 2001 semestre 1 hasta el periodo 2018, semestre 2, un total de 36 registros. Gráficamente se puede verificar que la serie de tiempo maneja una estacionariedad en su comportamiento durante ciertos segmentos de su temporalidad. Como se observa la variable ha tenido un crecimiento en la última de las dos décadas ilustradas, sin embargo en los últimos 3 años ha tenido un decrecimiento.

La serie de tiempo maneja un proceso estocástico con las siguientes características, que comprenden a la distribución de los datos. Como principal evidencia se tiene



que los datos de la serie de tiempo no siguen una distribución normal y manejan una asimetría de los datos.



Gráfica 5 Distribución de la serie de tiempo

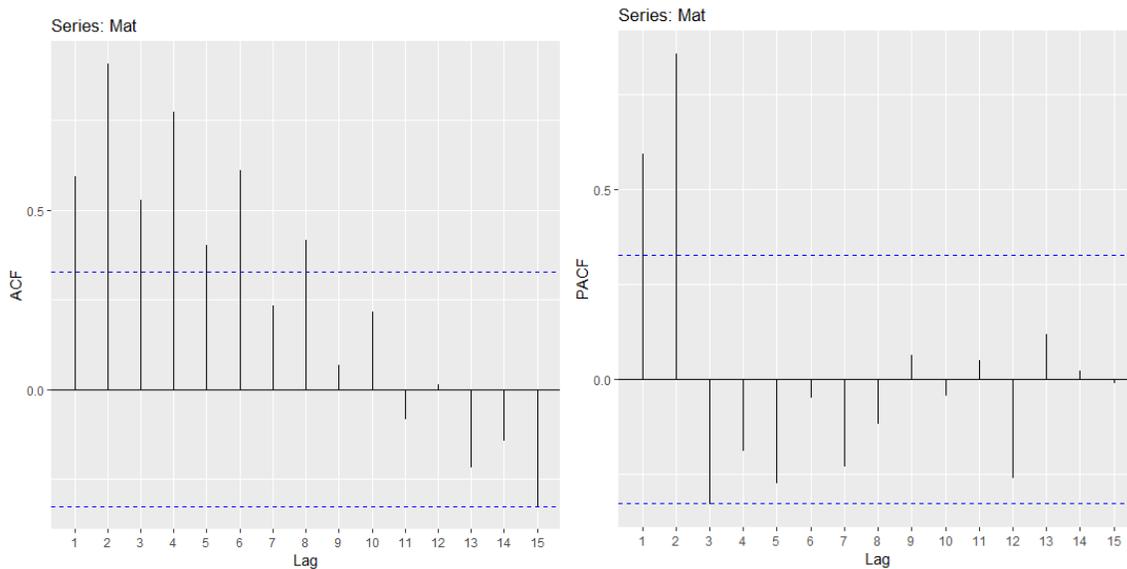
Fuente: Análisis en R

Sin embargo los resultados de la prueba Jarque Bera muestra que el p-valor es mayor que la significación de 0,05, por lo que los datos siguen una distribución normal.

X squared	P valor
2,518	0,2839

Tabla 1 Prueba Jarque Bera

Para verificar la estacionariedad de la serie de tiempo se trabajan con los gráficos de la función de auto correlación y auto correlación parcial, donde se tienen los siguientes comportamientos:



Gráfica 6 Diagrama ACF y PACF serie de tiempo

Se puede apreciar en esta gráfica como la serie no es estacionaria ya que el valor de la función de auto correlación no decae de manera exponencial a medida que aumentan los rezagos en el tiempo. Sin embargo con las pruebas de estacionariedad ADF, KPPS y PP tenemos los siguientes resultados:

Pruebas de estacionariedad

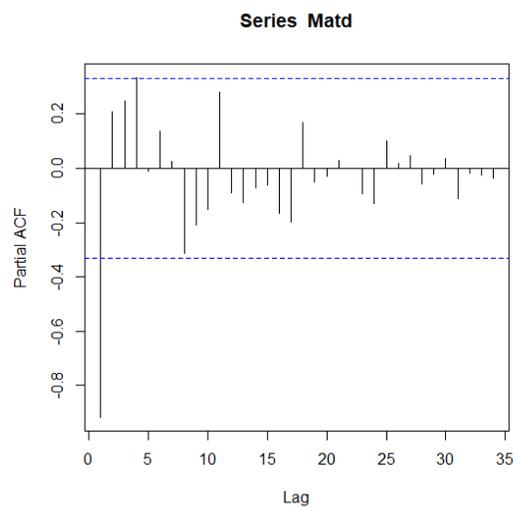
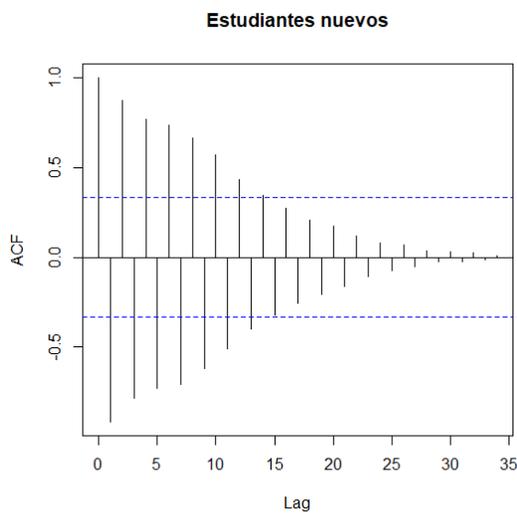
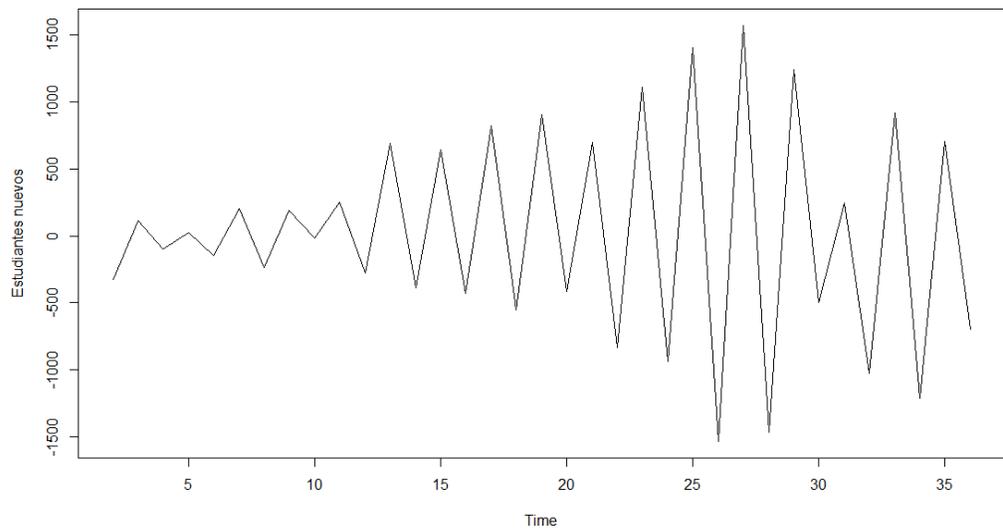
Prueba	Estadístico	P valor	Resultado
ADF	-0,25897	0.9868	Estacionaria
KPPS	0,81231	0,01	Estacionaria
PP	-53,279	0,01	Estacionaria

Tabla 2 Prueba de estacionariedad serie de tiempo

Teniendo esta contrariedad se propone de igual manera llevar a cabo transformación de los datos a través un nivel de diferenciación e identificar rezagos en la temporalidad de los datos.



La diferenciación de la serie de tiempo muestra una mayor estabilización del comportamiento de los datos



Gráfica 7 Análisis serie de tiempo diferenciada



De acuerdo a la función de auto correlación vemos como esta decae a medida que transcurre el tiempo, alternando sus valores, generando entonces una dependencia por valores muy cercanos, lo que da a entender que en la estimación del modelo se tendrá una presencia de coeficientes asociados

5.2 Estimación del modelo

El proceso de estimación del modelo que predice el comportamiento implicó un ajuste de del modelo a través de una diferenciación de la serie de tiempo así como con sus datos originales. Teniendo la siguiente evolución del modelo:

Modelo	BIC	AIC	RMSE	MAE
ARIMA(2,1,2)	502,4	510,22	260,75	2,5
ARIMA(3,1,1)	503,4	511,17	265,49	6,7
ARIMA (1,1,1)	504,13	508,79	286,1	45,9
ARIMA (6,0,2) fixed=c(0,NA,0,NA,0,NA,0,0,NA)	523	529,7	277,6	24,09
ARIMA (6,0,1), fixed=c(NA,0,NA,0,NA,0,0,NA)	583,98	591,9	264,15	3,72

Tabla 3 Comparación y ajuste de modelos

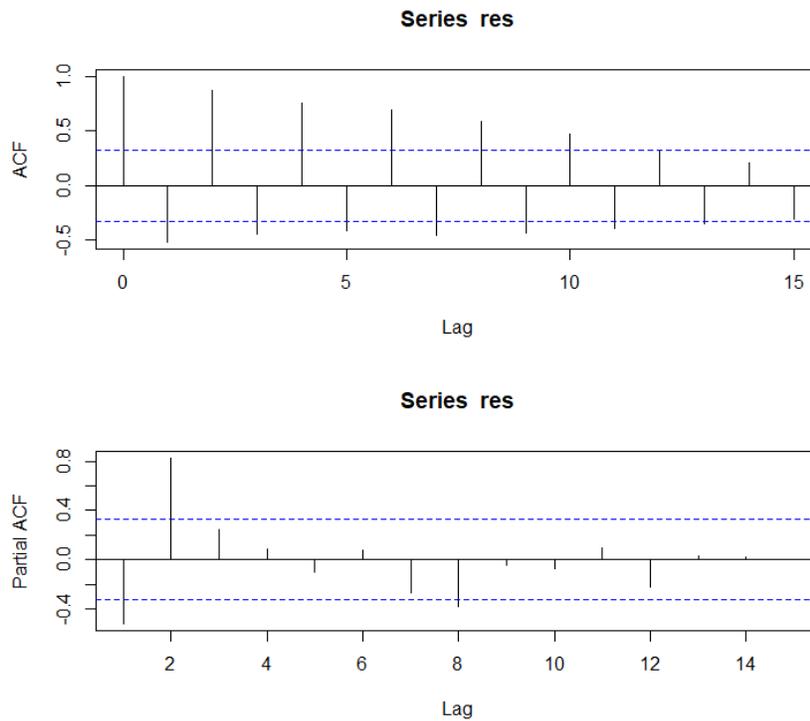
De acuerdo al ajuste realizado, el modelo que tiene un mayor ajuste al comportamiento esta representad por el siguiente modelo lineal

$$y_t = 0,5568y_{t-1} + 0,1914y_{t-3} - 0,1885y_{t-5} + 1398,58 + \varepsilon$$



5.3 Verificación del diagnóstico

El comportamiento de los residuos de este modelo se puede evidenciar en el siguiente gráficos que ilustran la función de autocorrelacion y autocorrelacion parcial, como se muestra en estos no existe una dependencia en rezagos anteriores lo que demuestra el comportamiento aleatorio de estos, especialmente en el grafico PACF.



Gráfica 8 Análisis ACF y PACF de residuos

De igual manera aplicando el test de Jarque Bera encontramos que estos siguen una distribución normal y la prueba Ljung Box muestra un comportamiento aleatorio y de independencia entre cada uno de los datos que componen a los residuos del modelo.

X squared	P valor
5,4081	0,06693

Tabla 4 Prueba Jarque Bera residuos

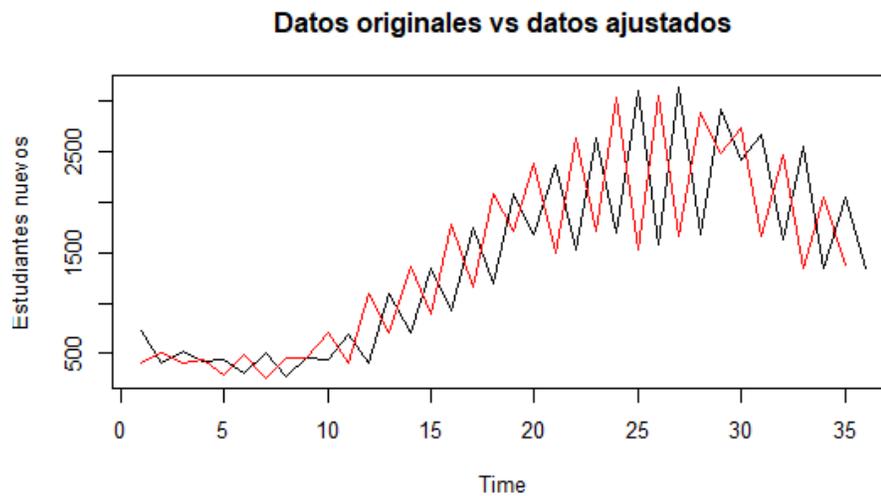


X squared	P valor
13,717	0,00002126

Tabla 5 Prueba Ljung Box

5.4 Pronostico

El modelo establecido maneja una estimación cercana al comportamiento real de las matriculas de nuevos estudiantes por semestre. En el anexo se puede ver los datos estimados en base al modelo ARIMA propuesto.



Gráfica 9 Datos originales Vs Datos ajustados

El pronóstico de nuevos estudiantes matriculados en el primero curso de los pregrados de la Fundación Universitaria Tecnológico Comfenalco de acuerdo al modelo arroja la siguiente información para los siguientes 3 años o 6 semestres:

Semestre	Pronostico
2019 1	2313,29
2019 2	1257,27



2020 1	2321,45
2020 2	1217,41
2021 1	2280,39
2021 2	1334,12

Tabla 6 Pronostico 2019 1- 2021 2 serie de tiempo

La información proyectada permite entonces tener un conocimiento para que la institución universitaria pueda planificar sus procesos a un largo plazo, identificando metas de procesos de matrículas y captación de nuevos usuarios de los servicios educativos semestre a semestre. La progresión de este pronóstico marca una estabilidad para los próximos años, lo cual puede estar impulsado por condiciones que se pueden dar igualmente a nivel nacional.



6. CONCLUSIONES

Con el planteamiento del modelo ARIMA se observa que la serie de tiempo tiene una principal dependencia de datos inmediatamente anterior al contemporáneo y disminuyendo esa dependencia a medida que se aleja hacia el pasado. El modelo no tiene una representación de un componente MA o de diferenciación de datos, debido a la presencia de estacionariedad en los datos originales de la serie de tiempo.

La predicción generada por el modelo ARIMA no tiene el ajuste indicado o idealmente esperado a la serie de tiempo, se requeriría de una mayor cantidad de información para dar una mayor robustez del modelo, sin embargo la predicción marcada por este modelo permite visualizar una aproximación del comportamiento de la serie de tiempo, la cual tendrá una estabilidad en la demanda teniendo en cuenta el comportamiento de los últimos años, teniendo una mayor participación de la demanda en el primer semestre de cada año y menor en el segundo de cada año.

Es importante que la estimación de este modelo sea retroalimentado con la inclusión de más información histórica y la intervención de variables exógenas que caracterizan la economía local y nacional. De esta manera poder tener un estimativo más confiable en la medida que pase el tiempo



7. REFERENCIAS

Abuín, J. R. (2007). Regresión lineal múltiple. Ideygdml-Estadística, Editor, 32.

Ardila Ariza, J. A., & Ávila Gómez, E. (2018). Construcción de una metodología empleando la herramienta R para estimar valores de los activos Bancolombia, Bogotá y occidente con modelos ARIMA.

Asteriou, D. (2002). Notas sobre Análisis de Series de Tiempo: Estacionariedad, Integración y Cointegración. Obtenido de Notes on Time Series Analysis: Stationarity, Integration and Cointegration

Castrillo-Rojas, D., & Rodríguez-Vargas, A. (2009). Procedimiento para la aplicación de pruebas de raíz unitaria. Banco Central de Costa Rica

Chong Fuentes, M. A. (2016). *Proyección de series de tiempo para el consumo de la energía eléctrica a clientes residenciales en Ecuador (Doctoral dissertation, Universidad de Concepción. Facultad de Ciencias Económicas y Administrativas).*

De Arce, R., & Mahía, R. (2003). Modelos Arima. *Programa CITUS: Técnicas de Variables Financieras.*

De Oliveira, E. S., De Almeida Dourado, J. D., & Mello, J. A. V. B. (2017). Aplicação de modelos de previsão de demanda em uma fábrica de embalagens plásticas. *Revista da Universidade Vale do Rio Verde*, 15(2), 354-373

Espinoza, M., Vicente, S., & Sempértegui Cañizares, E. R. (2017). Proyección a corto plazo de la demanda eléctrica de los alimentadores de distribución de la Empresa Eléctrica Regional Centro Sur CA (Master's thesis, Universidad del Azuay)

Forero-Lozano, S, Ballesteros-Ballesteros, V. y Toledo, J. (2018). Estadística de gradiente: una opción para realizar pruebas de hipótesis en escenarios de



tamaño de muestra pequeño. *Revista Internacional de Investigación de Ingeniería Aplicada*, 13 (23), 16368-16375

Guerra, D., Gutiérrez, M., & Silva, A. (2016). Generando espacios lúdicos para mejorar la participación artística de los estudiantes del ciclo 5 de la Institución Educativa. *Repository. Libertadores. Edu. Co*, 55. Obtenido de [https://repository.Libertadores. Edu. Co/bitstream/handle/11371/844/Guti% C3% a9rr ezmar% C3% adaelena. Pdf](https://repository.Libertadores.Edu.Co/bitstream/handle/11371/844/Guti%C3%A9rrez%20Silva%20A.pdf).

Mamani, R. P. P. (2018). Proyección de la demanda de Turismo Internacional en Puno: Un enfoque SARIMA. *Semestre Económico*, 5(2).

Molano Díaz, Y. A., & Correa Ramos, Y. (2016). Predicción del fraude con tarjetas para una entidad financiera a través del modelo arimax.

Munenaka, E., & Carmen, V. (2017). Pronóstico de demanda utilizando la metodología de Box-Jenkins

Nyoni, T. (2018). Box-Jenkins ARIMA approach to predicting net FDI inflows in Zimbabwe.

Nyoni, T., & Nathaniel, S. P. (2018). Modeling rates of inflation in Nigeria: an application of

Render, B., & Heizer, J. (2007). *Administración de la Producción*. México: Hugo Rivera Oliver

Rodríguez-Jaume, M. J., & Mora Catalá, R. (2001). Análisis de regresión múltiple. *Técnicas de Investigación Social II*.

Romero Moreno, J. S. (2019). Un acercamiento a los modelos ARIMAX y Lineal dinámico para estimar natalidad en Colombia en el periodo entre enero 2011 a diciembre 2017.

Sotomonte-Castro, J. E., Rodríguez-Rodríguez, C. C., Montenegro-Marín, C. E., Gaona-García, P. A., & Castellanos, J. G. (2016). Hacia la construcción de un



FUNDACIÓN UNIVERSITARIA
LOS LIBERTADORES

modelo predictivo de deserción académica basado en técnicas de minería de datos.
Revista científica, 3(26), 37-52

SNIES(2019).estadísticas de inscritos a población estudiantil 2001 a 2018.
Consultado de <https://www.mineducacion.gov.co/sistemasinfo/Informacion-a-la-mano/212400:Estadisticas>

Torres Romero, A., Cubillos Fonseca, J. S., & Camelo Sánchez, F. E. (2015).
Análisis de serie de tiempo de la ejecución presupuestal de Inversión Pública en
Colombia entre 1954 y 2013.

Torres, W. C. (2017). Proyección del Consumo Eléctrico Residencial de la
Región Lambayeque Mediante Análisis Univariante Aríma. Revista Científica
Pakamuros, 5(1), 12-12.