

UNIVERSIDAD NACIONAL DEL ALTIPLANO
FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA



**ANÁLISIS UNIVARIANTE PARA DESCRIBIR Y PROYECTAR LA
DEMANDA DE PACIENTES DEL HOSPITAL REGIONAL MANUEL
NUÑEZ BUTRON – PUNO, PERIODO 2008- 2017**

TESIS

PRESENTADA POR:

YERSON CUCHO HUANCA

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO ESTADÍSTICO E INFORMÁTICO

PUNO – PERÚ

2019

UNIVERSIDAD NACIONAL DEL ALTIPLANO
FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA

ANÁLISIS UNIVARIANTE PARA DESCRIBIR Y PROYECTAR LA DEMANDA DE PACIENTES DEL HOSPITAL REGIONAL MANUEL NUÑEZ BUTRON – PUNO, PERIODO 2008 - 2017

TESIS PRESENTADA POR:

YERSON CUCHO HUANCA

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO ESTADÍSTICO E INFORMÁTICO

APROBADA POR EL JURADO REVISOR CONFORMADO POR:



PRESIDENTE

:

.....

Dr. LEONEL COYLA IDME

PRIMER MIEMBRO

:

.....

M.Sc. ELQUI YEYE PARI CONDORI

SEGUNDO MIEMBRO

:

.....

Dr. JOSE PANFILO TITO LIPA

DIRECTOR / ASESOR

:

.....

Dr. EDGAR ELOY CARPIO VARGAS

ÁREA : Estadística

TEMA : Series de tiempo

FECHA DE SUSTENTACIÓN : 10 de julio del 2019

DEDICATORIA

Al divino creador por darme la vida junto a mí querida familia y guiarme día a día, lo que hace más liviano mi camino

Con mucho respeto y cariño dedicado a mi madre Luzmila Huanca Mamani, y a mi padre Darío Cucho Chacca, por su infinito amor, esfuerzo, comprensión, sacrificio y apoyo incondicional en mi formación humana y profesional.

A mis hermanos Hugo Cucho Huanca y Leonel Cucho Huanca, por su apoyo que siempre me dieron para salir adelante en mi vida personal y profesional

A mis compañeros y amigos de la Escuela Profesional de Ingeniería Estadística e Informática quienes con sus ánimos y apoyo moral y a los gratos recuerdos que siempre estarán grabados en mi mente

AGRADECIMIENTOS

A Dios, guía y compañero incondicional hasta en los momentos sesgados de mi vida.

A mis padres y familia en general, por brindarme su apoyo y motivación durante mi permanencia en la universidad.

A la Universidad Nacional del Altiplano y a la Facultad de Ingeniería Estadística e Informática, Escuela Profesional Estadística e Informática, por acogerme en sus aulas estos cinco años de formación.

A los Catedráticos de ña Escuela Profesional de Ingeniería Estadística e Informática, por compartir sus conocimientos con sus estudiantes y contribuir en la formación profesional, por absolver cada una de mis dudas, por su paciencia y calma en las sesiones de aprendizaje, mi cariño, respeto y admiración por cada uno de ellos.

A los docentes, que me brindaron sus conocimientos, me guiaron y apoyaron en la realización de la presente investigación.

Un agradecimiento muy grande también a los jurados Dr. Leonel Coyla Idme, M. Sc. Elqui Yeye Pari Condori, Dr. Jose Panfilo Tito Lipa y en especial a mi director de tesis Dr. Edgar Eloy Carpio Vargas que con su dirección y apoyo eficaz hicieron posible la conclusión de la presente tesis.

ÍNDICE GENERAL

ÍNDICE DE FIGURAS	7
ÍNDICE DE TABLAS	8
ÍNDICE DE ACRÓNIMOS	9
RESUMEN	10
ABSTRACT	11

CAPÍTULO I

INTRODUCCIÓN

1.1. FORMULACIÓN Y DEFINICIÓN DEL PROBLEMA.....	13
1.2. OBJETIVOS DE LA INVESTIGACIÓN	15
1.2.1 Objetivo general	15
1.2.2 Objetivos específicos	15
1.3. HIPÓTESIS	15

CAPÍTULO II

REVISIÓN DE LITERATURA

2.1 ANTECEDENTES DE LA INVESTIGACIÓN.....	16
2.2 BASE TEÓRICA	20
2.2.1. Procesos estocásticos	20
2.2.2. Pronósticos	20
2.2.3. Series de tiempo	21
2.2.4. Elementos estadísticos en el análisis de series de tiempo	23
2.2.5 Técnicas de predicción	23
2.2.6. Modelos de series temporales.....	24
2.2.7. Procesos estocásticos estacionarios	25

CAPÍTULO III**MATERIALES Y MÉTODOS**

3.1. POBLACIÓN.....	72
3.2. MUESTRA	72
3.3. UNIDAD MUESTRAL	72
3.4. METODO DE RECOLECCION DE DATOS	72
3.5. METODOLOGIA	73
3.6. METODO DE ANALISIS DE DATOS	73
3.6.1. Método de box – jenkins (teoría de wiener – kolmogorov).	73

CAPÍTULO IV**RESULTADOS Y DISCUSIÓN**

4.1. ANÁLISIS DE LA SERIE MENSUAL DEL NÚMERO DE PACIENTES ATENDIDOS EN LOS CONSULTORIOS EXTERNOS DEL HOSPITAL REGIONAL MANUEL NÚÑEZ BUTRÓN	78
4.1.1. Identificación del modelo	79
4.1.2. Fase de estimación del modelo identificado: arima (1.1.1)(0.1.1) ₁₂	84
4.1.3. Validación del modelo.....	85
4.1.4. Realización del pronóstico con el modelo	89
CONCLUSIONES	91
RECOMENDACIONES	91
REFERENCIAS BIBLIOGRÁFICAS	93
ANEXOS	95
Anexo 1: Datos originales del número de atenciones según meses del periodo 2008 – 2017	96
Anexo 2: primera diferencia de los datos del número de pacientes	98
Anexo 3: Residuo de los datos del número de pacientes	99

ÍNDICE DE FIGURAS

	Pág.
Figura 1: Coeficiente de autocorrelación y de autocorrelación parcial de los modelos AR(1) y AR(2).....	62
Figura 2: Coeficiente de autocorrelación y de autocorrelación parcial de los modelos MA(1) y MA(2).....	63
Figura 3: Coeficientes de autocorrelación y autocorrelación parcial de un modelo ARMA(1, 1).....	64
Figura 4: Metodología del enfoque Box – Jenkins.....	74
Figura 5: Evolución del número de pacientes atendidos, periodo 2008-2017.	79
Figura 6: Correlograma FAC y FACP de la serie del número de pacientes atendidos, periodo 2008-2017.....	81
Figura 7: Primera diferencia regular de la serie del número de pacientes atendidos, periodo 2008-2017.....	82
Figura 8: Correlograma FAC y FACP de la primera diferencia regular de la serie del número de pacientes atendidos, periodo 2008-2017.....	83
Figura 9: Residuales del modelo estimado de la serie del número de pacientes.	86
Figura 10: Correlograma FAC y FACP de los residuales del modelo de la serie del número de pacientes atendidos, periodo 2008-2017.....	86
Figura 11: Grafico Q para evaluar la normalidad de los residuales del modelo.	88
Figura 12: Predicción del número de pacientes atendidos para el año 2018.....	90

ÍNDICE DE TABLAS

	Pág.
Tabla 1: Serie del número de pacientes atendidos en consultorios externos, periodo 2008-2017.....	78
Tabla 2: Comparativo de modelos estimados de la serie del número pacientes atendidos, periodo 2008-2017.....	84
Tabla 3: Prueba Ljung-Box Q' de la serie del número de pacientes atendidos.....	87
Tabla 4: Pronóstico de la serie del número de pacientes.	89

ÍNDICE DE ACRÓNIMOS

ARIMA: acrónimo del inglés Autorregresive Moving Average

AR: Autorregresiva

MA: Media Móvil

ARMA: acrónimo del inglés AutoRegressive Moving Average Models

AIC: Criterio de Información Akaike

ε_t : Error aleatorio o residuo. También conocido como “Ruido Blanco”

Y_t : Variable Dependiente

W_t : Primera diferencia

d: Número de diferenciaciones

T: Número total de observaciones

RESUMEN

El presente trabajo de investigación se realizó en el Hospital Regional Manuel Núñez Butrón-Puno con el objetivo encontrar un buen modelo univariante. La metodología empleada para la serie de demanda de pacientes del Hospital es Box – Jenkins, la población en estudio es el número total de pacientes registrados mensualmente en los diferentes consultorios externos, la muestra está conformada por el número de pacientes registrados mensualmente en los consultorios externos del Hospital, los datos mensuales fueron obtenidos del libro de informes de los diferentes consultorios externos correspondientes a los periodos 2008 – 2017. El resultado obtenido de esta investigación, siguiendo los objetivos específicos. Un buen modelo univariante que se ajusta para describir y proyectar la demanda de pacientes del Hospital regional Manuel Núñez Butrón Puno es el modelo ARIMA. Los parámetros del modelo son (1.1.1)(0.1.1)₁₂, con ecuación:

$$\hat{Y}_t = -0.07 * Y_{t-1} - 0.99 * Y_{t-2} - 0.78 * Y_{t-3} + Y_{t-12} - Y_{t-13} + \varepsilon_t - \varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}.$$

Al validar dicho modelo, los residuos del modelo tienen un comportamiento de Ruido Blanco. Así mismo, se realizó las predicciones para el siguiente año 2018 con un nivel de confianza del 95% y el pronóstico revela una réplica con un ajuste bueno de la serie.

Palabras clave: Box-Jenkins, estimar, parámetros, predicción, univariante.

ABSTRACT

The present research work was carried out at the Manuel Núñez Butrón-Puno Regional Hospital with the following objective: to find a good univariate model. The methodology used for the patient demand series of the Hospital is Box - Jenkins, the population under study is the total number of patients registered monthly in the different outpatient clinics, the sample is made up of the number of patients registered monthly in the outpatient clinics of the Hospital, the monthly data were obtained from the book of reports of the different external offices corresponding to the periods 2008 - 2017. The result obtained from this investigation, following the specific objectives. A good univariate model that fits to describe and project the demand of patients from the Manuel Núñez Butrón Puno regional Hospital is the ARIMA model. The parameters of the model are (1.1.1) (0.1.1)₁₂, with equation:

$$\hat{Y}_t = -0.07 * Y_{t-1} - 0.99 * Y_{t-2} - 0.78 * Y_{t-3} + Y_{t-12} - Y_{t-13} + \varepsilon_t - \varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}.$$

When validating said model, the residuals of the model have a White Noise behavior. Likewise, predictions were made for the following year 2018 with a confidence level of 95% and the forecast reveals a replica with a good fit of the Series

Keywords: Box-Jenkins, estimate, parameters, prediction, univariate.

CAPÍTULO I

INTRODUCCIÓN

Una de las dificultades que se presentara en el futuro será sin duda, la falta de conocimiento del número de atenciones en consultorios externos, ya que no hay evidencia de trabajos de investigación que describan su comportamiento en un futuro. El Hospital Regional Manuel Núñez Butrón carece de modelos de predicción por lo que se puso énfasis en el siguiente trabajo de investigación con el único efecto de determinar dicho modelo con el cual el Hospital podrá hacer pronósticos, y estos ayudaran a prever recursos humanos, infraestructura, equipamiento, recursos tecnológicos y financieros que organizados adecuadamente, deben solucionar las necesidades de salud de la población puneña.

Utilizando modelos Box-Jenkins, se obtuvo un buen modelo de predicción. Por lo tanto el estudio de las series de tiempo, dentro de este los modelos univariantes constituye una necesidad ya que en la actualidad son muchos los campos en los que se pueden aplicar estos conocimientos, de manera que permite tomar acertadas decisiones, puesto que se apoyan en técnicas para la aceptación o rechazo de hipótesis y poder actuar a partir de las series observadas, de esta manera predecir el comportamiento de la variable en estudio, es así que permitirá proyectar la demanda de pacientes del Hospital Regional Manuel Núñez Butrón, para el año 2018.

La presente investigación consta de siete capítulos donde:

En el capítulo I, se explica los fundamentos para la realización de la tesis, así como planteamiento del problema y definición del problema, justificación objetivos e hipótesis.

En el capítulo II, se describe el marco teórico y presenta los diversos conceptos necesarios para el correcto entendimiento de la tesis, que consiste en: antecedentes de la investigación, base teórica.

En el capítulo III, se describe los materiales y métodos para el modelo de predicción, que se utilizaron en la presente investigación.

En el capítulo IV, se realiza el análisis, se muestra los resultados del mejor modelo univariante para proyectar la demanda de pacientes en el Hospital Regional Manuel Núñez Butrón.

En el capítulo V, se presenta las conclusiones que se obtuvieron en el análisis de los datos

En el capítulo VI, se mencionan las recomendación y sugerencias sobre el modelo univariante de Box-Jenkins.

En el capítulo VII, se muestra las referencias para un mejor uso de toda la información presentada

1.1. FORMULACIÓN Y DEFINICIÓN DEL PROBLEMA

El Hospital Regional Manuel Núñez Butrón - Puno, es una institución a cargo del estado (pública) de cierto grado de complejidad, de mucha referencia a nivel de la región de Puno, orientado exclusivamente a la atención de toda la población de Puno por ende cuenta con profesionales en las diferentes áreas de servicio al público, de igual manera cabe recordar que la población tiene un continuo crecimiento esto implica que dicha institución debe mejorar la oferta y la calidad de la atención en beneficio de la población del departamento de Puno, con cierto énfasis en los grupos más vulnerables (población

rural), la misma que está en busca de la mejora continua de calidad de servicio y de la fiabilidad del mismo.

El Hospital Regional Manuel Núñez Butrón - Puno y como casi la mayoría de establecimientos de salud, realiza evaluaciones mensuales , trimestrales, semestrales y anuales con la finalidad de obtener un determinado resultado de los servicios brindados, y estos son enviados a la DIRESA (Dirección Regional de Salud) - Puno durante el mes de trabajo con la finalidad de evaluar y programar el presupuesto para cada área para el siguiente año en curso de acuerdo al cumplimiento de las metas planteadas.

Otro punto del problema es, con el propósito de mejorar la oferta de servicios de salud aumenta el crecimiento desordenado de la oferta de servicios creando una imprecisión en la articulación de servicios e ineficacia del sistema de referencia, desorden administrativo, siendo necesario realizar las estimaciones a futuro; esto conlleva a que el nosocomio prevea los servicios y las evaluaciones que llegan a ocurrir en el transcurso de los años, para luego obtener resultados muy garantizados.

La razón del trabajo, es conseguir modelos uniecuacionales de series de tiempo (Técnica BOX JENKINS), para la variable del número de pacientes atendidos en el HRMNB –Puno, correspondiente a los periodos del año 2008 hasta el año 2017, que permitan realizar pronósticos adecuados para periodos cortos de tiempo en función a hechos pasados correlacionados, frente a los aspectos mencionados sobre el problema surge la siguiente interrogante:

¿Cuál es el modelo univariante que mejor describe y proyecta la demanda de pacientes del Hospital Regional Manuel Núñez Butrón – Puno en el periodo 2008 - 2017?

1.2. OBJETIVOS DE LA INVESTIGACIÓN

1.2.1 Objetivo general

Determinar un buen modelo univariante para proyectar la demanda de pacientes del Hospital Regional Manuel Núñez Butrón - Puno, en el periodo 2008 - 2017.

1.2.2 Objetivos específicos

- Determinar el modelo utilizando la tendencia de la serie para el número de pacientes en el Hospital Regional Manuel Núñez Butrón - Puno
- Determinar los parámetros del modelo que mejor se ajusta para el número de pacientes en el Hospital Regional Manuel Núñez Butrón - Puno
- Validar el modelo que mejor se ajusta para el número de pacientes en el Hospital Regional Manuel Núñez Butrón - Puno.
- Proyectar la demanda de pacientes en el Hospital Regional Manuel Núñez Butrón-Puno.

1.3. HIPÓTESIS

El modelo univariante no integrado de Box-Jenkins se ajusta mejor a la serie de número de pacientes del Hospital Regional Manuel Núñez Butrón - Puno, periodo 2008 - 2017.

CAPÍTULO II

REVISIÓN DE LITERATURA

2.1 ANTECEDENTES DE LA INVESTIGACIÓN

Bedoya (2018) “Modelamiento univariado del número de defunciones infantiles producidas por infecciones respiratorias agudas, a través de la metodología Box-Jenkins, Puno 2008-2016” (Tesis de pregrado) de la Escuela Profesional de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano Puno, Perú. Cuyo objetivo general es: Determinar un modelo univariado del número de defunciones infantiles producidas por infecciones respiratorias agudas, a través de la metodología Box-Jenkins. Llegando a la siguiente conclusión: El modelo de pronóstico univariado integrado seleccionado que mejor se ajusta para predecir el número de defunciones infantiles producidas por infecciones respiratorias agudas, en la población del distrito de Puno es ARIMA (0,1,1)(0,1,1), cuya ecuación de pronóstico estimada es:

$$\hat{Y}_t = Y_{t-1} + Y_{t-12} - Y_{t-13} - 0.96796e_{t-1} - 0.60690e_{t-12} + 0.58745e_{t-13}$$

Valdez (2015) ”Modelo de predicción mensual del número de intervenciones quirúrgicas más frecuentes en el Hospital regional Manuel Núñez butrón - Puno, 2006 - 2015-I” (Tesis de pregrado) de la Escuela Profesional de Ingeniería Estadística e Informática de la Universidad Nacional del altiplano, Puno, Perú. Cuyo objetivo general es: Determinar el modelo de predicción mensual que mejor ajuste para predecir el número de intervenciones quirúrgicas más frecuentes en el Hospital Regional Manuel Núñez Butrón - Puno en el periodo 2006 -2015-I. Llegando a la siguiente conclusión: Los modelos de predicción mensual que mejor se ajustan para decidir y predecir el comportamiento de la serie de tiempo del número de intervenciones quirúrgicas más

frecuentes son: cesárea $SARIMA(2.1.0)(0.1.0)_{12}$, aborto $SARIMA(1.1.1)(0.1.1)_{12}$, laparotomía $SARIMA(1.1.2)(0.1.0)_{12}$, colecistectomía $SARIMA(0.1.1)(1.1.0)_{12}$, apendicetomía $SARIMA(2.1.0)(0.1.0)_{12}$.

Segunda.- El mejor modelo útil para la predicción son: cesárea $\hat{Y}_t = -0.53 * Y_{t-1} - 0.48 * Y_{t-2} + Y_{t-12} - Y_{t-13} + \varepsilon_t - \varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}$, aborto $\hat{Y}_t = -0.18Y_{t-1} + Y_{t-12} - Y_{t-13} + \varepsilon_t + \varepsilon_{t-1} + 0.28\varepsilon_{t-12} - 0.28\varepsilon_{t-13}$, laparotomía $\hat{Y}_t = 0.79Y_{t-1} + Y_{t-12} - Y_{t-13} + \varepsilon_t - 1.6\varepsilon_{t-1} - 0.6\varepsilon_{t-2} - \varepsilon_{t-12} + \varepsilon_{t-13}$, colecistectomía $\hat{Y}_t = Y_{t-1} - 0.18Y_{t-12} + 0.18Y_{t-13} + \varepsilon_t + \varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}$, apendicetomía $\hat{Y}_t = -0.5Y_{t-1} - 0.5Y_{t-2} + Y_{t-12} - Y_{t-13} + \varepsilon_t - \varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}$.

Quispe (2008) “Modelo de predicción de los nacimientos y defunciones del distrito de Juliaca-2008” (Tesis de pregrado) de la Escuela profesional de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano, Puno, Perú. Cuyo objetivo general es: Determinar modelos univariantes para la predicción del número de nacimientos y el número de defunciones en la población del Distrito de Juliaca, para el año 2008. Llegando a la siguiente conclusión: Primera.- El modelo de pronóstico univariante integrado ARIMA (2,1,0)x(2,1,3)₁₂ que mejor se ajusta para predecir el número de nacimientos para el año 2008 en el distrito de Juliaca es:

$$\begin{aligned}
\widehat{Y}_t = & Y_{t-1} + Y_{t-12} - Y_{t-13} - 0.80Y_{t-1} + 0.380Y_{t-2} + 0.80Y_{t-13} - 0.80Y_{t-14} \\
& - 0.30Y_{t-2} + 0.30Y_{t-13} + 0.30Y_{t-14} - 0.30Y_{t-15} + 0.80Y_{t-12} \\
& - 0.80Y_{t-13} - 0.80Y_{t-24} + 0.80Y_{t-25} - 0.76Y_{t-24} + 0.76Y_{t-25} \\
& + 0.76Y_{t-36} - 0.76Y_{t-37} + 0.64Y_{t-13} - 0.64Y_{t-14} - 0.64Y_{t-25} \\
& + 0.64Y_{t-26} - 0.61Y_{t-25} + 0.61Y_{t-26} + 0.61Y_{t-37} - 0.61Y_{t-38} \\
& + 0.24Y_{t-14} - 0.24Y_{t-25} - 0.24Y_{t-26} + 0.24Y_{t-27} - 0.23Y_{t-26} \\
& + 0.23Y_{t-37} + 0.23Y_{t-38} - 0.23Y_{t-39} - 2.11\varepsilon_{t-13} + 1.69\varepsilon_{t-24} \\
& - 0.51\varepsilon_{t-36}
\end{aligned}$$

Y el modelo de pronóstico univariante integrado ARIMA (0,1,1)x(2,0,1)₁₂ que mejor se ajusta para predecir el número de defunciones para el año 2008 en el distrito de Juliaca es: $\widehat{Y}_t = Y_{t-1} + 0.74Y_{t-12} + 0.74Y_{t-13} + 0.53Y_{t-24} - 0.53Y_{t-24} - 0.53Y_{t-25} - 0.74\varepsilon_{t-1} + 1.15\varepsilon_{t-12} - 0.85\varepsilon_{t-13}$

Segunda.- se obtuvieron las predicciones mensuales para el año 2008 para las series de los nacimientos y las defunciones de la población de la provincia de Juliaca con los modelos estimados.

Melo (2016) “Modelo de predicción mensual de mortalidad general intrahospitalaria en el Hospital Regional Manuel Núñez Butrón, Puno 2008-2016 I” (Tesis de Pregrado) de la Escuela Profesional de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano, Puno, Perú. Cuyo objetivo general es: Determinar el mejor modelo univariante de predicción mensual del número de mortalidad general intrahospitalaria en el Hospital Regional Manuel Núñez Butrón, Puno 2008-2016 I. Llegando a la siguiente conclusión: Primero.- El modelo de predicción mensual que mejor se ajusta para decidir y predecir el comportamiento de la serie de tiempo del número de morbilidad intrahospitalaria es ARIMA (2, 1,0) (0, 1,1)₁₂. Segundo, al determinar la

predicción en los modelos alcanzados para la serie de número de mortalidad intrahospitalaria del Hospital Regional Manuel Núñez Butrón-Puno, revelan una réplica bastante buena y proporcionan una alternativa eficaz para describir y predecir el comportamiento futuro del número de mortalidad intrahospitalaria, el mejor modelo útil para la predicción es: $\hat{Y}_t = Y_{t-1} * Y_{t-2} + Y_{t-2} - (0.58)Y_{t-12} - (0.29)Y_{t-13} + \varepsilon_t + 0.999 - \varepsilon_{t-12} + \varepsilon_{t-13}$, este modelo fue elegido por los criterios de Akaike y Schwarz, y por comparación de datos y ajuste con la serie original.

Merlín (2015) “Modelo univariante de pronóstico del número de unidades de transfusión de sangre en el Hospital Regional Manuel Núñez Butrón, Puno periodo 2006-2015 I” (Tesis de Pregrado) de la Escuela Profesional de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano, Puno, Perú. Cuyo objetivo general es: Determinar el modelo de predicción mensual que mejor se ajusta para realizar el pronóstico del número de unidades de transfusión de sangre en el Hospital Regional Manuel Núñez Butrón, Puno. Llegando a la siguiente conclusión: Los modelos de predicción mensual que mejor se ajusta para decidir y predecir el comportamiento de la serie de tiempo del número de unidades de transfusión de sangre es: ARIMA (0,1,1)(1,1,0)12, cuya ecuación de pronóstico es: $\hat{Y}_t = Y_{t-1} + Y_{t-2} - (0.36)Y_{t-12} - (0.36)Y_{t-13} + \varepsilon_t + (0.89)\varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}$

Quispe (2015) “Modelo univariante para el consumo de energía eléctrica doméstica en el distrito de Ayaviri – Electro Puno, periodo 2004-2013” (Tesis de Pregrado) de la Escuela Profesional de Ingeniería Estadística e Informática de la Universidad Nacional del Altiplano Puno, Perú. Cuyo objetivo general es: Determinar el modelo univariante de ajuste que nos permita describir y predecir el Consumo de Energía Eléctrica en el distrito de Ayaviri – Electro Puno, periodo 2004-2013. Llegando a la siguiente conclusión: El mejor modelo univariante que nos permite describir y predecir

el comportamiento del consumo de Energía Eléctrica doméstica en el Distrito de Ayaviri -Puno, es el siguiente modelo: ARIMA (3, 1,1) (1, 0,1)

$$\hat{Y}_t = Y_{t-1} - 1.0247Y_{t-1} - Y_{t-2} + 0.3257Y_{t-2} - Y_{t-3} - 0.1801Y_{t-3} - Y_{t-4} + 0.7447Y_{t-12} - Y_{t-13} + 0.7630Y_{t-13} - Y_{t-14} + 0.2425Y_{t-14} - Y_{t-15} + 0.1341Y_{t-15} - Y_{t-16} - 0.6113\varepsilon_{t-12} - 0.917\varepsilon_{t-1} - 0.561\varepsilon_{t-13} + \varepsilon_t$$

2.2 BASE TEÓRICA

2.2.1. Procesos estocásticos

Uriel (1985) un proceso estocástico se define como una familia de variables aleatorias que corresponden a momentos sucesivos del tiempo. Será por $Y(t, u)$, donde t es el tiempo y u es la variable aleatoria. La determinación de las características de un proceso estocástico puede hacerse mediante dos formas alternativas, bien a partir de funciones de distribución conjunta o bien a partir de los momentos. En general, para un conjunto finito de valores del tiempo, se obtiene una función de distribución conjunta, así para $t_1, t_2 \dots t_n$ la función de distribución conjunta sería: $F[Y(t_1), Y(t_2) \dots Y(t_n)]$. Si se dispone de un conjunto finito de estas variables $t_1, t_2 \dots t_n$ se dice que esta sucesión de observaciones forma una serie temporal.

Alberto (2007) es una secuencia de variables aleatorias, ordenadas y equidistantes cronológicamente, referidas a un (proceso univariante o escalar) o a varias (proceso multivariante o vectorial) características de una unidad observable en diferentes momentos.

2.2.2. Pronósticos

Uriel (1985) los pronósticos son predicciones de lo que puede suceder o esperar, son premisas o suposiciones básicas en que se basan la planeación y la toma de decisiones.

El propósito del pronóstico consiste en reducir el margen de incertidumbre, haciendo el mejor uso de la información que se tiene para guiar las actividades de la empresa hacia el cumplimiento de sus metas y objetivos. Los pronósticos se basan en el uso de datos anteriores de una variable para predecir su desempeño futuro. A este respecto, los datos anteriores se dan generalmente en la forma de series de tiempo. Una hipótesis básica en la aplicación de las técnicas de pronóstico es que el desempeño de los datos anteriores continuara ocurriendo en el futuro inmediato.

2.2.3. Series de tiempo

Anderson (1985) define como un conjunto de observaciones en el tiempo según una característica cuantitativa de un fenómeno individual en diferentes momentos del tiempo, en el cual las observaciones son realizadas.

Grass y Ferrer (1994) una serie de tiempo consta de datos que se reúnen, registran u observan sobre incrementos sucesivos de tiempo. En el análisis de series de tiempo de datos, una tendencia inmediata consiste en intentar explicar o contabilizar el comportamiento de las series. La tendencia de una serie de tiempo es el componente de largo plazo que representa el crecimiento o disminución en la serie sobre un periodo amplio, el componente cíclico es la fluctuación en forma de onda alrededor de la tendencia, el componente estacional es un patrón de cambio que se repite a sí mismo año tras año, el componente aleatorio mide la variabilidad de las series de tiempo después de retirar los otros componentes

Aznar, García y Martín (1994) manifiesta que la serie de tiempo es una información básica de la evolución de las variables en el tiempo, los modelos de Box-Jenkins constituyen un conjunto de procedimientos para el tratamiento de series de tiempo.

Hanke (1996) es un conjunto de observaciones ordenadas según una característica cuantitativa de un fenómeno individual en diferentes momentos del tiempo, en el cuál las observaciones son realizadas

Pindyck y Rubinfeld (2003) define que la serie temporal al modelo que se supone anda sobre las relaciones causales que afecta a las variables en estudio. Consiste en examinar el comportamiento de la serie en el pasado, para luego inferir el comportamiento futuro.

La planificación racional exige prever los sucesos del futuro que probablemente vayan a ocurrir. La previsión, a su vez, se suele basar en lo que ha ocurrido en el pasado. Se tiene pues un nuevo tipo de inferencia estadística que se hace acerca del futuro de alguna variable o compuesto de variables basándose en sucesos pasados. La técnica más importante para hacer inferencias sobre el futuro con base en lo ocurrido en el pasado, es el análisis de series de tiempo.

Alberto (2007) una serie temporal (o simplemente una serie) es una secuencia de N observaciones (datos) ordenadas y equidistantes cronológicamente sobre una característica (serie univariante o serie escalar) o sobre varias características (serie multivariante o vectorial) de una unidad observable en diferentes momentos.

Guerrero (2009) el nombre de “serie de tiempo” no es del todo apropiado para denotar un conjunto de datos registrados de manera ordenada respecto al tiempo, pues es particular el término serie se utiliza en matemáticas para nombrar a una suma infinita de valores de una variable. Una terminología más apropiada para referirse al conjunto de datos que interesa podría ser el de sucesiones cronológicas; sin embargo, se continuara haciendo mención a series de tiempo, debido a que esta es la terminología más usual y común.

2.2.4. Elementos estadísticos en el análisis de series de tiempo

Debido a que las series de tiempo constan de datos numéricos, es natural usar la herramienta de la estadística para describirlas y analizarlas, así como ocurre con cualquier otro conjunto de información numérica.

Anderson (1985) el análisis de series de tiempo está dedicado al estudio de series, por lo general, los datos de dichas series son independientes pero que están correlacionados; se puede decir que existe una relación entre observaciones contiguas.

Idea central

Uriel (1985) si el objetivo es explicar el valor que toma, en un momento determinado del tiempo, un fenómeno económico que muestra dependencia temporal, un procedimiento factible consiste en recoger información sobre su evolución a lo largo del tiempo, y explotar el patrón de regularidad que muestran los datos.

Para construir un modelo de series de tiempo, lo único que se necesita es la información muestral de la variable a analizar.

2.2.5 Técnicas de predicción

Hanke (1996) el objetivo de las técnicas de predicción es obtener estimaciones o pronósticos de valores futuros de una serie temporal a partir de la información histórica contenida en la serie observada hasta el momento actual. Estas técnicas no requieren la especificación de los factores que determinan el comportamiento de la variable, sino que se basan únicamente en la modelización del comportamiento sistemático de la serie.

Hanke (1996) las predicciones se basan con el uso de datos anteriores de una variable para predecir su desempeño futuro. A este respecto, los datos anteriores se presentan, generalmente en la forma de series de tiempo. Una hipótesis básica, en la aplicación de las técnicas de predicción, fue el desempeño de los datos anteriores continúan ocurriendo en el futuro inmediato. Evidencias empíricas indican que este supuesto es válido en muchas situaciones reales, sobre todo cuando las series de tiempo representan una larga historia de las variables analizadas.

Ventajas y desventajas

Ventajas

- Son más sencillos de estimar.
- Con niveles de desagregación temporal elevados (datos mensuales) es mucho más fácil construir un modelo de series temporales

Desventajas

- A menudo no se dispone de los datos de las variables exógenas (por ejemplo expectativas).
- Dificultades en el marco del método de regresión para la estimación con variables retardadas (especialmente con variable endógena retardada).
- Predicción: como predecir los valores de las variables exógenas.

2.2.6. Modelos de series temporales

Variables Temporales: Variables que se observan a lo largo del tiempo. Y_t indica la variable Y en el momento t .

Serie Temporal: Conjunto de T observaciones, una observación por cada una de las variables Y_1, Y_2, \dots, Y_t Sinónimo: serie cronológica.

A las observaciones de una variable temporal se denomina realizaciones. Son los resultados de un proceso estocástico.

Uriel (1985) en gran parte de las variables temporales propias del análisis económico se detecta un patrón de comportamiento en el tiempo. Las series presentan movimientos sistemáticos (repetitivos). En la evolución en el tiempo se observa una realidad. Esta regularidad de las series temporales es, en general, estocástica y no determinística. Es decir, esa regularidad no admite una formulación funcional determinista, sino que es función de variables aleatorias. Estimado tal regularidad, el analista construye un mecanismo explicativo que recibe el nombre de modelo.

2.2.7. Procesos estocásticos estacionarios

Se denomina proceso estocástico a la sucesión infinita de variables aleatorias ordenadas. Y_1, Y_2, \dots, Y_t

Si se dispone de un conjunto finito de estas variables, Y_1, Y_2, \dots, Y_t , se dice que esta sucesión de observación (realizaciones) forma una serie temporal.

Ruidos blancos

Un tipo especial de proceso estocástico es el denominado ruido blanco.

Una variable a_t se denomina “ruido blanco” si se cumple las siguientes condiciones:

$$i. \quad E(\varepsilon_t) = 0 \quad \forall t$$

$$ii. \quad V(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2 \quad \forall t$$

$$iii. \quad COV(\varepsilon_t, \varepsilon_s) = 0 \quad \forall t \neq s$$

Uriel (1985) esto es, la variable x_t presenta una esperanza constante en el tiempo, e igual a 0; varianza constante e Incorrelación temporal (autocorrelación nula). Si se agrega la condición de que la variable Y_t se distribuye normal, la condición (iii) de Incorrelación implica independencia.

Supondremos, salvo que se exprese lo contrario, que la distribución de probabilidad asociada a un ruido blanco es normal. Ello significa que las 3 condiciones anteriores pueden sintetizarse en la expresión:

$$\varepsilon_t \sim \forall t$$

Uriel (1985) el término ruido ha sido tomado de la teoría de la comunicación. En esta disciplina, un ruido corresponde a perturbaciones aleatorias que vienen en función de la señal enviada, que es diferente de la señal recibida. Se denomina ruido blanco a una perturbación aleatoria con las características antes planteadas, por analogía con la distribución de energía continua en luz blanca de un cuerpo incandescente.

Utilización de las series de tiempo

Una de las herramientas estadísticas para uso en pronósticos de sucesos futuros que están, en alguna forma, entrelazados con la economía es el análisis de tiempo. Los fabricantes están en extremo interesados en los ciclos de altibajos de la propia economía así de las economías extranjeras de modo que puedan predecir mejor la demanda de sus

productos, que a su vez impacta sus niveles de inventarios, requerimientos de personal, flujos de efectivo y casi todas las demás actividades de negocios dentro de la empresa.

Hanke (1996) “Los científicos políticos están interesados en el uso de análisis de series de tiempo para estudiar los patrones de cambio del gasto de gobierno en defensa y programas de bienestar social. Es obvio que estas tendencias tienen un gran impacto en el futuro de industrias complejas”.

Análisis de las series de tiempo

Anderson (1985) el análisis de series de tiempo está dedicado al estudio de series; por lo general, los datos de dichas series son independientes pero están correlacionados; se puede decir que existe una relación entre observaciones contiguas.

Es el análisis de una secuencia de medidas hechas a intervalos específicos. El tiempo es usualmente la dimensión dominante de los datos. Sirven para establecer la efectividad de medidas que afectan a grupos poblacionales teniendo en cuenta las variaciones naturales que puede haber en el tiempo. Son muy comunes en la evaluación de leyes en la población. Permiten una visión parcial de la relación causa efecto, pero no pueden extrapolar los hallazgos de la población a individuos específicos.

El análisis de series de tiempo consiste en una descripción (general matemática) de los movimientos y componentes presentes.

De acuerdo a Chatfield (1978) son varios los objetivos por los cuales se desea analizar una serie de tiempo:

Descripción: Al tener una serie de tiempo, el primer paso en el análisis es graficar los datos y obtener medidas descriptivas simples de las propiedades principales de la serie.

Explicación: Cuando las observaciones son tomadas sobre dos o más variables, es posible usar la variación en una serie para explicar la variación en las otras series.

Predicción: Dada una serie de tiempo se intenta predecir los valores futuros de la serie. Este es el objetivo más frecuente en el análisis de series de tiempo.

Control: Si una serie de tiempo se genera por mediciones de calidad de un proceso, el objetivo del análisis puede ser el control del proceso.

Series de tiempo discretas

Guerrero (2003) menciona que: “Una serie de tiempo es la sucesión de observaciones generales por un proceso estocástico, cuyo conjunto índice se toma en relación al tiempo. Por tanto, la inferencia que se realice será acerca de las características del proceso estocástico generador de la serie de observaciones”.

Proceso estocástico y estacionariedad

Aznar y Trivez (1993) explica que: “los modelos autorregresivos o de medias móviles tienen un proceso estocástico”.

Un proceso estocástico es una sucesión de variables aleatorias Y_t ordenadas, pudiendo tomar t cualquier valor entre $-\infty$ y ∞ . Por ejemplo, la siguiente sucesión de variables aleatorias puede ser considerada como proceso estocástico.

$$Y_{-5}, Y_{-4}, Y_{-3}, Y_{-2} \dots Y_3, Y_4$$

El subíndice t no tiene, en principio, ninguna interpretación a priori, aunque si hablamos de proceso estocástico en el contexto del análisis de series temporales este subíndice representará el paso del tiempo.

Cada una de las variables Y_t que configuran un proceso estocástico tendrán su propia función de distribución con sus correspondientes momentos. Así mismo, cada par de esas variables tendrán su correspondiente función de distribución conjunta y sus funciones de distribución marginales. Esto mismo ocurrirá, ya no para cada par de variables, sino para conjuntos más amplios de las mismas. De esta forma, para caracterizar un proceso estocástico deberíamos especificar las funciones de distribución conjunta de cualquier conjunto de variables:

$$Y_{t_1}, Y_{t_2}, Y_{t_3} \dots Y_{t_m}$$

Cualesquiera que fueran los valores de $(t_1, t_2, t_3 \dots t_m)$ y cualquiera que fuera el valor de m ; por ejemplo:

$$Y_1, Y_2, Y_3 = (t_1 = 1 \text{ y } m = 3)$$

$$Y_3, Y_4, Y_5, Y_6 = (t_1 = 3 \text{ y } m = 4)$$

Habitualmente, conocer esas funciones de distribución resulta complejo de forma que, para caracterizar un proceso estocástico, basta con especificar la media y la varianza para cada Y_t y la covarianza para variables referidas a distintos valores de t :

$$E(Y_t) = \mu_t$$

$$\sigma_t^2 = \text{var}(y_t) = E[y_t - \mu_t]^2$$

$$y_t = \text{cov}(Y_y, Y_s) = [(y_t - \mu_t)(y_s - \mu_s)]$$

Universidad Autónoma de Madrid (2015) las distribuciones de probabilidad podrían no estar completamente caracterizadas en algunas de las variables, los momentos podrían no coincidir incluso no existir para alguna de las variables aleatorias, lo mismo puede ocurrir con las distribuciones conjuntas o marginales. Sin embargo, de todos los

tipos de procesos estocásticos posibles, nos interesan especialmente dos de ellos a los que la estadística ha dado nombres precisos.

Ruido blanco

Es una sucesión de variables aleatorias (proceso estocástico) con esperanza (Media) cero, varianza constante e independientes para distintos valores de t (covarianza nula).

Uriel (1985) el ruido blanco es una señal aleatoria (proceso estocástico) que se caracteriza porque sus valores de señal en dos instantes de tiempo diferentes no guardan correlación estadística. Como consecuencia de ello, su densidad espectral de potencia (PSD, Power Spectral Density) es una constante, su grafica plana. Esto significa que la señal contiene todas las frecuencias y todas ellas tienen la misma potencia. Igual fenómeno ocurre con la luz blanca, lo que motiva la denominación.

Proceso estocástico estacionario

Decimos que un proceso estocástico es estacionario si las funciones de distribución conjuntas son invariantes con respecto a un desplazamiento en el tiempo (variación de t). Es decir, considerando que $t, t+1, t+2, \dots, t+k$ reflejan períodos sucesivos:

$$F(Y_t, Y_{t+1}, \dots, Y_{t+k}) = F(Y_{t+m}, Y_{t+1+m}, \dots, Y_{t+k+m})$$

Para cualquier t, k y m ; por ejemplo:

$$F(Y_1, Y_2, \dots, Y_6) = F(Y_{10}, Y_{11}, \dots, Y_{15})$$

Donde $t=1, k=5, m=9$

$$F(Y_3, Y_4, Y_5) = F(Y_7, Y_8, Y_9)$$

Donde $t=3, k=2, m=4$

Universidad Autónoma de Madrid (2015) se conoce como estacionariedad en sentido estricto o fuerte y puede relajarse sustancialmente utilizando la denominada estacionariedad en sentido amplio o débil. Decimos que un proceso estocástico es débilmente estacionario si:

- Las esperanzas matemáticas de las variables aleatorias no dependen del tiempo, son constantes:

- $E[Y_t] = E[Y_{t+m}] \quad \forall m$

- Las varianzas tampoco dependen del tiempo y son finitas:

- $Var[Y_t] = Var[Y_{t+m}] \neq \infty \quad \forall m$

- Las covarianzas entre dos variables aleatorias del proceso correspondientes a períodos distintos de tiempo:

- $Cov[Y_t, Y_s] = Cov[Y_{t+m}, Y_{s+m}] \quad \forall m$

De esta última condición se desprende que, si un fenómeno es estacionario, sus variables pueden estar relacionadas linealmente entre sí, pero de forma que la relación entre dos variables sólo depende de la distancia temporal k transcurrida entre ellas.

Lógicamente, la estacionariedad en sentido estricto garantiza la estacionariedad en sentido amplio pero no al revés.

Una vez introducido el concepto genérico de proceso estocástico puede decirse que una serie temporal cualquiera es, en realidad, una muestra, una realización concreta con unos valores concretos de un proceso estocástico teórico, real. El análisis de series que vamos a estudiar tratará, a partir de los datos de una serie temporal, inferir las

características de la estructura probabilística subyacente, del verdadero proceso estocástico.

Modelos autorregresivos

Universidad Autónoma de Madrid (2015) definimos un modelo como autorregresivo si la variable endógena de un período t es explicada por las observaciones de ella misma correspondientes a períodos anteriores añadiéndose, como en los modelos estructurales, un término de error. En el caso de procesos estacionarios con distribución normal, la teoría estadística de los procesos estocásticos dice que, bajo determinadas condiciones previas, toda Y_t puede expresarse como una combinación lineal de sus valores pasados (parte sistemática) más un término de error (innovación).

Los modelos autorregresivos se abrevian con la palabra AR tras la que se indica el orden del modelo: AR (1), AR (2),...etc. El orden del modelo expresa el número de observaciones retrasadas de las series temporales analizadas que intervienen en la ecuación. Así, por ejemplo, un modelo AR (1) tendría la siguiente expresión:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + e_t$$

El término de error de los modelos de este tipo se denomina generalmente ruido blanco cuando cumple las tres hipótesis básicas tradicionales mencionadas al principio del texto:

- i. Media nula
- ii. Varianza constante
- iii. Covarianza nula entre errores correspondientes a observaciones diferentes

La expresión genérica de un modelo autorregresivo, no ya de un AR (1) sino de un AR (p) sería la siguiente:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t$$

Pudiéndose escribir de forma abreviada como:

$$\phi_p(L)Y_t = \phi_0 + e_t$$

Donde $\phi_p(L)$ es lo que se conoce como operador polinomial de retardos:

$$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$$

Y donde, a su vez, el término L es lo que se conoce como operador retardo tal que, aplicado al valor de una variable en t, dé como resultado el valor de esa misma variable en t-1:

$$LY_t = Y_{t-1}$$

Y aplicado sucesivamente p veces retarda el valor en p períodos

$$L^p Y_t = Y_{t-p}$$

Normalmente, se suele trabajar con modelos autorregresivos de órdenes bajos: AR (1) o AR (2), o bien con órdenes coincidentes con la periodicidad de los datos de la serie analizada (si es trimestral AR (4), si es mensual AR (12)).

Modelo de medias móviles

Un modelo de los denominados de medias móviles es aquel que explica el valor de una determinada variable en un período t en función de un término independiente.

Las medias móviles son indicadores que allanan o suavizan, en mayor o menor medida, de tal manera que eliminan determinadas fluctuaciones, sean a corto, medio o largo plazo.

Estos modelos se denotan normalmente con las siglas MA, seguidos, como en el caso de los modelos autorregresivos, del orden entre paréntesis. Así, un modelo con q términos de error MA (q) respondería a la siguiente expresión:

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

Que de nuevo puede abreviarse utilizando el polinomio de retardos (como en caso de los modelos AR):

$$Y_t = \theta_q(L)\varepsilon_t + \mu$$

Al igual que en el caso de los modelos autorregresivos, el orden de los modelos de medias móviles suele ser bajo MA(1), MA(2) o corresponderse con la periodicidad de los datos analizados MA(4), para series trimestrales, o MA(12) para series mensuales.

Interpretación de un modelo de medias móviles

Universidad Autónoma de Madrid (2015) así como un modelo autorregresivo es intuitivamente sencillo de comprender, la formulación de un modelo de medias móviles resulta sorprendente para el no iniciado. ¿Qué significa que una variable aleatoria se explique en función de los errores cometidos en períodos precedentes?, ¿De dónde proceden esos errores?, ¿Cuál es la justificación de un modelo de este tipo?

En realidad, un modelo de medias móviles puede obtenerse a partir de un modelo autorregresivo sin más que realizar sucesivas sustituciones.

Efectivamente, un modelo AR (1), sin término independiente, tiene la expresión:

$$Y_t = \phi Y_{t-1} + e_t$$

Si consideramos t-1 en lugar de t el modelo sería en este caso:

$$Y_{t-1} = \phi Y_{t-2} + e_{t-1}$$

Y sustituyendo queda:

$$Y_t = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 Y_{t-2}$$

Si ahora sustituimos Y_{t-2} por su expresión autorregresivo y así sucesivamente llegamos a un modelo del tipo:

$$Y_t = \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 Y_{t-2} + \phi^3 \varepsilon_{t-3} + \dots + \phi^j \varepsilon_{t-j} + \dots$$

Que es la expresión, sin término independiente, de un modelo de medias móviles como el planteado anteriormente. En realidad, de forma estricta, el paso de un modelo a otro debería realizarse al contrario (de un MA a un AR) utilizando el teorema general de descomposición de Wold.

Condiciones y raíces unitaria para los modelos AR Y MA

Universidad Autónoma de Madrid (2015) hemos dicho anteriormente que, bajo condiciones generales, todo proceso estocástico estacionario se prestaba a una especificación tipo AR (p) y en consecuencia podía expresarse también como un MA (q). Es ahora el momento de especificar lo que antes hemos llamado "condiciones generales" y examinar en qué casos es posible la realización de un proceso AR o MA para representar un proceso estocástico estacionario.

Para que un proceso estocástico estacionario admita una formulación del tipo que aquí estudiaremos han de cumplirse dos condiciones accesorias:

- El proceso no debe ser anticipante (hipótesis de recursividad temporal); lo que quiere decir que los valores de una variable en un momento t no dependerán de los que esta misma tome en $t + j$, siendo j cualquier valor superior a cero.
- El proceso ha de ser invertible; lo que supone que la correlación entre una variable y su pasado va reduciéndose a medida que nos alejamos más en el tiempo del momento para el que estamos considerando dicha correlación (proceso ergódico). La explicación intuitiva de esta situación derivaría de que si el especificáramos una variable en función de ciertos coeficientes que nos determinen su correlación con los valores pasados de ella misma, los valores de dichos coeficientes deberían ser necesariamente inferiores a uno, porque si no el proceso de infinitos números sería "explosivo".

La estacionariedad de las series temporales en la realidad

Universidad Autónoma de Madrid (2015) Queda clara que la aproximación a los procesos estocásticos con modelos AR o MA está restringida, en términos generales, a aquellos procesos estocásticos que cumplan, al menos de forma débil, la restricción de estacionariedad. Cuando, en la realidad, queremos inferir a partir de una serie temporal (muestra) la estructura del proceso estocástico mediante modelos AR o MA, debemos cubrir dos etapas:

- Asegurarnos de que la serie temporal, como muestra del proceso estocástico, es estacionaria y, si no lo es.

- Transformar la serie temporal original de forma que la nueva serie transformada si lo sea.

A. ¿cómo verificamos si la serie a analizar es estacionaria en media? ¿Cómo lograr que lo sea? ¿Filtrado de la serie original?

Universidad Autónoma de Madrid (2015) Para resolver la primera cuestión existen diversos métodos de aproximación y, de entre ellos. Podríamos subdividir la serie temporal en varios períodos la misma longitud y calcular su media. El proceso sería estacionario en el caso en que dichos estadísticos fueran prácticamente iguales para todos los sub períodos analizados. En la mayoría de los casos un simple gráfico sirve para observar si existe o no una clara tendencia y por tanto si la serie es estacionaria o no.

Habitualmente cuando una serie muestra tendencia, se subdivide dicha serie en dos componentes: una primera, la estimación de dicha tendencia, y, la segunda, el residuo o error que se comete cuando se utiliza dicha tendencia como valor estimado de la serie original.

$$y_t = T_t + r_t$$

Una vez estimada la tendencia, aproximada con una regresión lineal, parabólica, exponencial que sea más conveniente; trabajaremos con la serie del residuo, que entonces no mostrara tendencia y podremos decir que es estacionaria en media. Es sobre este residuo sobre el que llevaremos a cabo todo el proceso descrito como metodología de identificación ARIMA, sumando finalmente el valor de la tendencia estimada si queremos dar resultados de estimación de la serie original. Es decir:

La identificación del proceso ARIMA se hará sobre esta serie del residuo $r_t = y_t - T_t$, estimada previamente la tendencia del modo más adecuado.

Para obtener valores estimados de la serie original, sumaremos el componente tendencial al valor estimado del residuo mediante el modelo ARIMA $y_t = T_t + r_t$.

A este procedimiento se le conoce con el nombre de filtrado de la tendencia de la serie. Por supuesto, existen muy variadas formas de aplicar un filtro, siendo la que aquí hemos enunciado la más sencilla.

B. ¿Cómo se comprueba si una serie es estacionaria en varianza? Orden de integración

Sin duda alguna, el test más habitual a la hora de determinar la estacionariedad de una serie temporal, consiste en la aplicación del conocido como test de Dickey–Fuller (Test DF) o Dickey-Fuller Ampliado (Test ADF) (Universidad Autónoma de Madrid, 2015). Éste es un contraste de “No estacionariedad” ya que la hipótesis nula es precisamente la presencia de una raíz unitaria en el proceso generador de datos de la serie analizada.

Vamos a suponer inicialmente, como modelo de partida para el análisis de una determinada serie Y_t , el de un proceso estacionario autorregresivo de orden uno:

$$Y_t = \varphi_1 y_{t-1} + \varepsilon_t \quad (\text{Ec. 1})$$

Frente a este modelo se plantea, como hipótesis nula H_0 , el modelo alternativo de un paseo aleatorio no estacionario del tipo:

$$Y_t = y_{t-1} + \varepsilon_t \quad (\text{Ec. 2})$$

Se trata por tanto de contrastar si el coeficiente φ_1 es igual a la unidad o distinto de ella.

Sin embargo, para contrastar la nulidad del coeficiente φ_1 , no podemos utilizar el contraste “t” habitual sobre una estimación por MCO del primer modelo. La razón es que la hipótesis nula que habitualmente se contrasta y, a partir de la cual se deriva la expresión y propiedades del test “t”, es la de nulidad del parámetro ($\varphi_1=0$) de la (Ec.2), sin embargo, en nuestro caso, necesitaríamos contrastar $H_0: \varphi_1=1$. Si la hipótesis nula fuera cierta, la varianza de Y_t no sería estacionaria sino que crecería con los valores de “t” según la expresión de la varianza de un paseo aleatorio con deriva:

$$Var(y_t) = t\sigma_\varepsilon^2$$

La estimación de a_1 en $y_t = \varphi_1 y_{t-1} + \varepsilon$ será siempre consistente sin embargo, su distribución variará según los valores que tome la estimación. La distribución de probabilidad asintótica del estimador de MCO del modelo AR (1) presenta una “discontinuidad” cuando $a_1=1$ y, como sustituto, deberán utilizarse las distribuciones derivadas de forma empírica mediante un procedimiento de Montecarlo realizado por Dickey.

En la práctica, se obtiene restando a uno y otro lado el término y_{t-1} :

$$y_t - y_{t-1} = a_0 + a_1 y_{t-1} - y_{t-1} + \varepsilon_t$$

$$\Delta y_t = a_0 + (a_1 - 1)y_{t-1} + \varepsilon_t \quad (\text{Ec. 3})$$

$$\Delta y_t = a_0 + \gamma \cdot y_{t-1} + \varepsilon_t$$

Por tanto, la hipótesis nula inicial para la (Ec. 2), se transforma ahora en $H_0: \gamma = 0$ frente a $H_1: \gamma < 0$. Decir que γ es nulo es lo mismo que decir que $\varphi_1=1$, o sea, que existe una raíz unitaria, decir que es menor que cero equivale a decir que a_1 es menor que la unidad (proceso autorregresivo estacionario) ó sea, no se considera el caso de procesos autorregresivos explosivos en que $\varphi_1 > 1$.

El procedimiento básico para la aplicación simple del test DF es, a partir de aquí, aparentemente sencillo. Se estima el modelo propuesto y se calcula el valor estimado de la “t” del parámetro analizado. Una vez calculado se compara con el valor empírico de referencia obtenido con las tablas de Dickey y Fuller o de MacKinnon. Si el valor estimado para γ es inferior en valor absoluto al tabulado dado un determinado nivel de confianza, admitiremos la hipótesis nula, o sea, la presencia de una raíz unitaria.

Connell, Bowerman y koehler (2007) el modelo expuesto hasta el momento es el más simple posible, pero cabe que el modelo más adecuado a la realidad incluya otros términos, como una constante y/o una tendencia.

Propusieron, entre otros autores, seguir un proceso en etapas a fin de garantizar el éxito en la elección del modelo de referencia en el mayor número de ocasiones:

- En primer lugar se estimaría el modelo menos restringido (con término constante y tendencia determinista).
- Dado que el principal error de esta táctica inicial consistiría en la escasa potencia del contraste para el rechazo de la hipótesis nula por inclusión de variables irrelevantes, si los valores críticos indican rechazo (ausencia de raíz unitaria), terminaríamos el procedimiento.
- En el caso de no rechazarse la hipótesis nula de presencia de una raíz unitaria, es decir, en el caso en que admitamos la presencia de una raíz unitaria ($H_0: \gamma = 0$) pasaríamos ahora a examinar la significatividad del parámetro tendencial determinista a_2 . Dado que, en este punto, estaríamos bajo la hipótesis ya admitida de que $\gamma = 0$, utilizaríamos el valor de referencia de $\tau_{\beta\tau}$ e incluso, para mayor seguridad, también el contraste conjunto $\Phi_3 (a_2 = \gamma = 0)$.

- Si el término tendencial resulta significativo ($a_2 \neq 0$) contrastaremos de nuevo la presencia de una raíz unitaria ($H_0: \gamma = 0$) pero utilizando entonces las tablas de una normal estandarizada. Sea cual sea el resultado del test con las nuevas tablas finalizaríamos aquí el contraste admitiendo o rechazando la presencia de una raíz unitaria.
- Si el término tendencial es no significativo, deberá replantearse el modelo inicialmente estimado pasándose a examinar otro con término constante pero sin esta tendencia determinista. Con este modelo se vuelve a analizar la presencia de una raíz unitaria ($\gamma = 0$).
- En el caso en que, nuevamente, se sostenga la presencia de una raíz unitaria, se contrastará entonces la adecuación del término independiente a_0 bien con el contraste $\tau_{\alpha\mu}$, bien con ϕ_1 . Si el término independiente resulta significativo usamos de nuevo las tablas de una normal para contrastar la presencia de la raíz unitaria, concluyendo de nuevo aquí el contraste.
- Sólo si entonces la constante a_0 es no significativa se utiliza el modelo más simple como modelo de referencia contrastándose, de nuevo, la presencia de raíz unitaria. En este caso, no tiene cabida el uso de la distribución normal estandarizada.

Universidad Autónoma de Madrid (2015) está claro que lo expuesto hasta este momento permite contrastar la presencia de una o más raíces unitarias en una determinada serie temporal para la que se supone un proceso AR (1). Sin embargo, muchas serie temporales se ajustan más adecuadamente a procesos autorregresivos de orden superior AR (2) o AR (3). No parece, por tanto, muy correcto, contrastar la presencia de una o más raíces unitarias utilizando siempre la estructura de un modelo AR (1) ya que las raíces unitarias pueden aparecer también en estructuras más complejas. Este problema da lugar

a lo que se conoce como test de raíces unitarias de Dickey-Fuller Ampliado (DFA): si se quiere contrastar la presencia de una raíz unitaria en una serie que sigue un proceso AR (p), deberá aplicarse el procedimiento expuesto para el caso simple AR (1), pero suponiendo ahora del modelo:

$$\Delta y_t = \varphi_0 + \gamma \cdot y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t$$

Dónde:

$$\gamma = -(1 - \sum_{i=1}^p e_i) \quad y \quad \beta_i = \sum_{j=1}^p e_j$$

Modelo arima (p, d, q).

Es un modelo que permite describir un valor como una función lineal de datos anteriores y errores debidos al azar. Se analiza sobre una serie estacionaria.

Los modelos de promedio móvil autorregresivo integrado (ARIMA: Autorregresive Integrated moving-average) son una clase especializada de técnicas de filtración que ignoran por completo a las variables independientes en la formulación de pronósticos. Estos modelos son dispositivos altamente refinados de ajuste de curvas que utilizan valores reales y anteriores de la variable dependiente, para producir pronósticos precisos de corto plazo.

En 1970, Box y Jenkins desarrollan un cuerpo metodológico destinado a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo juega una papel fundamental. Podemos decir que la consideración exclusiva de los valores pasados de una determinada variable para explicar su evolución presente y futura supone, al mismo tiempo una ventaja y un inconveniente:

- La ventaja radica en el hecho de no necesitar distintas series de datos (distintas variables) referida al mismo periodo de tiempo (característica común a todos los modelos univariantes) y, al mismo tiempo, ahorramos la identificación y especificación del modelo en el sentido de la econometría tradicional.
- El inconveniente es que, al renunciar a la inclusión de un conjunto más amplio de variables explicativas, no atenderemos a las relaciones que sin duda existen entre casi todas las variables económicas perdiendo capacidad de análisis de tiempo que renunciamos, implícitamente, al estudio teórico previo del fenómeno y de su indudable utilidad.

Los modelos ARIMA (p,d,q) constituyen una clase particular de procesos no estacionarios y se define como:

$$W_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} - \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \dots - \omega_q \varepsilon_{t-q}$$

Donde:

$$W_t = Y_t - Y_{t-1} \quad : \text{Primera diferencia}$$

d : Número de diferenciaciones

La mayor parte de las series económicas corresponden a procesos no estacionarios. Así, si se desea obtener un tratamiento de las series basado en el análisis de series de tiempo (modelo ARIMA), es necesario discutir mecanismos de transformación de las series a procesos estacionarios.

Funciones de un proceso estocástico estacionario

Definido un proceso estocástico como estacionario (ya sea de forma débil o fuerte), ya se ha comentado que si cumple las condiciones en sentido estricto, también

cumple las condiciones en sentido débil. Siendo así, el proceso estaría perfectamente definido si conociéramos su media constante (μ), su varianza constante (σ) y la covarianza entre cada par de momentos diferentes en el tiempo.

Dicho esto:

- **La función de autocovarianza** vendrá definida por los distintos valores que tomaría dicha covarianza cuando cambiamos el lapso temporal entre las observaciones de la serie que manejamos.

Analíticamente, se podría expresar como:

$$\gamma_j = Cov(y_t, y_{t-j}) = E[(y_t - \mu)(y_{t-j} - \mu)]$$

Donde, evidentemente, cuando el valor de "j" es cero, tendríamos la varianza de la función:

$$\gamma_0 = Cov(y_t, y_{t-0}) = E[(y_t - \mu)^2] = \sigma^2$$

- **La función de autocorrelación** se define igualmente como:

$$\rho_j = \frac{Cov(y_t, y_{t-j})}{\sqrt{var(y_t)}\sqrt{var(y_{t-j})}}$$

Como nos encontramos ante un proceso definido como estacionario, la varianza es constante, por lo que podemos escribir:

$$\rho_j = \frac{\gamma_j}{\gamma_0}$$

Función de autocorrelación parcial

Con el fin de tener en cuenta los valores de correlación entre dos variables aleatorias separadas entre sí "j" períodos y en función de los valores intermedios entre ellas.

Es decir:

$$\Pi_j = \text{corr}(y_t, y_{t-1}, y_{t-2}, \dots, y_{t-j+1})$$

Si planteamos las mejores predicciones de y_t e y_{t-j} como los resultantes de plantear Mínimos Cuadrados Ordinarios (MCO) siendo el primero de ellos del siguiente modo:

$$\hat{y}_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_{j-1} y_{t-j+1}$$

Se puede escribir la función de autocorrelación parcial, si la media es nula, como:

$$\Pi_j = \frac{\text{cov}[(y_t - \hat{y}_t)(y_{t-j} - \hat{y}_{t-j})]}{\sqrt{\text{var}(y_t - \hat{y}_t)}\sqrt{\text{var}(y_{t-j} - \hat{y}_{t-j})}}$$

Pudiendo demostrarse que:

$$\Pi_j = \frac{\rho_j - \alpha_1 \rho_{j-1} - \alpha_2 \rho_{j-2} - \dots - \alpha_{j-1} \rho_1}{1 - \alpha_1 \rho_1 - \alpha_2 \rho_2 - \dots - \alpha_{j-1} \rho_{j-1}}$$

Aplicación de las funciones a muestras concretas

Universidad Autónoma de Madrid (2015) en este apartado se pretender especificar estimaciones de los valores que caracterizan el proceso estacionario. Habrá que estimar la media (μ), para lo que usaremos la media muestral; la varianza (γ_0) y la función de

autocovarianza (γ_j), para lo que emplearemos la fórmula de la covarianza muestral) y la función de autocorrelación (ρ_j).

- **Media muestral**

Como ya se ha dicho, el estimador $\mu = E(y_t)$ será la media muestral:

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t$$

Dicho estimador cumplirá dos propiedades:

- i. Insesgadez.- la esperanza de la media de la serie será igual a μ .

Esto se demuestra ya que:

$$E(\bar{y}) = \frac{1}{T} \sum_{t=1}^T E(y_t) = \frac{1}{T} \cdot T \cdot \mu = \mu$$

- ii. Consistencia.- es decir, la varianza se anula cuando ampliamos la muestra a la población siendo el estimador insesgado, propiedad que se cumplirá siempre que se dé la siguiente condición, que no desarrollamos:

- **Función de autocovarianza muestral.**

El estimador de γ_j se obtendrá según la siguiente expresión:

$$C_j = \frac{1}{T-j} \sum_{t=j+1}^T (y_t - \bar{y})(y_{t-j} - \bar{y})$$

Que, a pesar de ser sesgado, dicho sesgo será determinable y cada vez más reducido según se aumente la muestra.

- **Función de autocorrelación muestral**

Para su cálculo se recurrirá al cociente de funciones de autocovarianza del siguiente modo:

$$r_j = \frac{\gamma_j}{\gamma_0} = \frac{C_j}{C_0}$$

- **El estimador para la función de autocorrelación parcial**

Para a emplear se calculará según el método recursivo de Durbin del siguiente modo:

$$\hat{\phi}_{11} = r_1$$

$$\hat{\phi}_{i+1,j+1} = \frac{r_{j+1} - \sum_{i=1}^j \hat{\phi}_{ji} r_{j+1-i}}{1 - \sum_{i=1}^j \hat{\phi}_{ji} r_i}$$

$$\hat{\phi}_{ji}, i = \hat{\phi}_{ji} - \hat{\phi}_{j+i,i+1} \hat{\phi}_{j,j+1+i} / i=1 \dots j$$

Proceso estocástico estacionario lineal discreto

Universidad Autónoma de Madrid (2015) vamos a definir un nuevo caso especial de un proceso estocástico que nos permita luego intentar encontrar algo parecido en la realidad y que sea fácilmente identificable

Se conoce por proceso estocástico estacionario lineal y discreto a aquel que puede expresarse como:

$$y_t = \mu + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots$$

Dónde:

- Es lineal porque puede escribirse como combinación lineal de los errores.

- e_t es ruido blanco (esperanza y covarianza nulas y varianza constante).
- Es discreto porque los lapsos temporales considerados son uniformes (no hay saltos temporales distintos entre las variables consideradas).

El siguiente paso será especificar las condiciones que nos aseguran que este proceso es estacionario, es decir que tiene media y varianza constantes y que su covarianza sólo varía cuando lo hace el espacio temporal que separa las observaciones empleadas para calcularla. Para ver estas condiciones, calcularemos los momentos de primer y segundo orden asegurando la estacionariedad en sentido débil.

- **Media constante**

$$E(y_t) = E(\mu + e_t + \theta_1 e_{t-1} + \dots) = \mu + \sum_{i=0}^{\infty} \theta_i E(e_t) \quad / \quad \theta_0 = 1$$

$$y \sum_{i=1}^{\infty} \theta_i = k < \infty$$

Es decir la media será constante en la medida en que exista la segunda parte del sumando que, al quedar multiplicada por la esperanza del "ruido blanco" será finalmente cero, y la media quedará igual a μ .

- **Varianza constante**

$$\gamma_0 = E(y_t - \mu)^2 = E(\mu - \mu^2 + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots)^2 =$$

$$E(e_t^2 + \theta_1^2 e_{t-1}^2 + \dots + 2\theta_1 e_t e_{t-1} + 2\theta_2 e_t e_{t-2} + \dots) =$$

$$\sigma_e^2 (1 + \theta_1^2 + \theta_2^2 + \dots + 0) =$$

$$\sigma_e^2 \sum_{i=0}^{\infty} \theta_i^2 \quad / \quad \theta_0 = 1$$

Siendo entonces condición necesaria para que la varianza exista que el último sumatorio sea calculable (converja).

- **Covarianza constante**

$$y_j = E[(y_t - \mu)(y_{t-j} - \mu)] =$$

$$E[(e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots)(e_{t-j} + \theta_1 e_{t-j-1} + \theta_2 e_{t-j-2} + \dots)] =$$

$$\theta_j E((e_{t-j}^2 + \theta_1 \theta_{j-1} E(e_{t-j-1}^2) + \theta_2 \theta_{j+2} E(e_{t-j-2}^2)) + \dots) =$$

$$\sigma_e^2 (\theta_j + \theta_1 \theta_{j+1} + \theta_2 \theta_{j+2} + \dots) = \sigma_e^2 \sum_{i=0}^{\infty} \theta_i \theta_{i+j}$$

Luego el proceso será estacionario en la medida en que se cumplan estas tres condiciones:

$$\sum_{i=0}^{\infty} \theta_i E(e_t) \quad / \theta_0 = 1 \quad y \quad \sum_{i=1}^{\infty} \theta_i = K < \infty$$

$$\sum_{i=0}^{\infty} \theta_i^2 < \infty \quad / \theta_0 = 1$$

$$\sum_{i=0}^{\infty} \theta_i \theta_{i+j} < \infty$$

- **Ventajas**

Una vez definido este proceso particular, vamos a ver resumidamente sus ventajas respecto a no contar con él.

En principio, si quisiéramos definir un proceso estocástico en general, tendríamos, al menos, que definir sus momentos de primer y segundo orden, para lo cual sería necesario estimar T varianzas, T esperanzas y $(T^2 - T)/2$ covarianzas, lo que nos es imposible si sólo contamos con T datos.

Si el proceso fuera estacionario, ya sólo tendríamos que estimar una esperanza y una varianza (media y una varianza constantes) y (T-1) covarianzas ($cov(y_t, y_{t-1})$, $cov(y_t, y_{t-2}), \dots$) ; En total $1+1+(T-1)=T+1$ parámetros, lo que tampoco es posible.

Si estamos ante un proceso estocástico estacionario lineal discreto, sólo necesitaremos contar con:

$$T > p + q + 2$$

Siendo "p" y "q" los órdenes de los retardos de los modelos autorregresivos y de medias móviles que ya hemos definido anteriormente.

Modelos MA (1)

Una vez tenemos definidas las ventajas de contar con un proceso estocástico estacionario lineal y discreto, y que podemos calcular las funciones de autocovarianza y autocorrelación, puede resultar interesante ver que valores toman éstas en aquellos casos sencillos que luego nos permitan comprobar si series generales pueden tener un comportamiento similar, simplemente acudiendo a la comparación de sus correlogramas (de la función de autocorrelación total y parcial).

El primer caso a analizar será el modelo de medias móviles de orden uno, que se define como:

$$y_t = \mu + e_t + \theta_1 e_{t-1}$$

Este modelo también se puede escribir en función del operador retardo, ya comentado, del siguiente modo:

$$y_t = \mu + \theta(L)e_t$$

$$\theta(L) = 1 - \theta_1 L$$

Se dice que un modelo MA (q) es invertible en la medida en que se pueda escribir como un proceso autorregresivo de orden infinito. Para que esta circunstancia pueda darse, será condición necesaria que las raíces de:

$$1 - \theta(L) = 0$$

Caigan fuera del círculo unitario, lo que se cumplirá siempre que $|\theta| < 1$.

Hanke y Reitsh (2006) esta situación proviene de la conversión del modelo de medias móviles en modelo autorregresivo. Si escribimos el MA (1) como:

$$y_t = \mu + e_t + \theta_1 e_{t-1}$$

Podemos hacer sucesivas sustituciones hasta llegar al modelo autorregresivo:

$$e_{t-1} = y_{t-1} - \mu - \theta_1 e_{t-2}$$

$$y_t = \mu + (y_{t-1} - \mu - \theta_1 e_{t-2}) + e_t$$

$$\theta_1 y_{t-1} - \theta_1^2 y_{t-2} + \dots + \mu(1 + \theta_1 + \theta_1^2 + \theta_1^3 + \dots + \theta_1^{n-1}) + e_t - \theta_1^n e_{t-n}$$

Si $n \rightarrow \infty$:

$$y_t = \theta_1 y_{t-1} + \theta_1^2 y_{t-2} + \dots + \mu \frac{1}{1 - \theta_1} + e_t$$

Donde es necesario que $|\theta| < 1$ para que la progresión geométrica que se produce sobre los parámetros θ sea calculable y no explosiva.

El siguiente paso, una vez definida la condición de invertibilidad, es definir las funciones que se han descrito para los procesos estocásticos en general para el caso del modelo MA (1).

- **Esperanza**

$$E(y_t) = E(\mu + e_t + \theta_1 e_{t-1}) = E(\mu) + E(e_t) + \theta_1 E(e_{t-1}) = \mu$$

- **Varianza**

$$var(y_t) = E(y_t - \mu)^2 = E(\mu + e_t + \theta_1 e_{t-1} - \mu)^2$$

$$E(e_t^2 + \theta_1^2 e_{t-1}^2 + 2\theta_1 e_t e_{t-1}) =$$

$$= \sigma_e^2 + \theta_1^2 \sigma_e^2 + 0$$

- **Función de autocovarianza**

$$\gamma_1 = cov(y_t, y_{t-1}) = E((y_t - \mu)(y_{t-1} - \mu)) =$$

$$= E((e_t + \theta_1 e_{t-1})(e_{t-1} + \theta_1 e_{t-2}))$$

$$E(e_t e_{t-1} + \theta_1 e_t e_{t-2} + \theta_1 e_{t-1} e_{t-1} + \theta_1^2 e_{t-1} e_{t-2})$$

$$\theta_1 E(e_{t-1})^2 = \theta_1 \sigma_e^2$$

$$\gamma_j = cov(y_t, y_{t-j}) = E((y_t - \mu)(y_{t-j} - \mu)) =$$

$$E((e_t + \theta_1 e_{t-1})(e_{t-j} + \theta_1 e_{t-j-1})) =$$

$$E(e_t e_{t-j} + \theta_1 e_{t-1} e_{t-j} + \theta_1 e_t e_{t-j-1} + \theta_1^2 e_{t-1} e_{t-j-1})$$

$$E(e_t e_{t-j} + \theta_1 e_{t-1} e_{t-j} + \theta_1 e_t e_{t-j-1} + \theta_1^2 e_{t-1} e_{t-j-1}) = 0 = \gamma_j$$

Luego la función de autocovarianza se anula para valores de "j" mayores que uno y es una fracción de la varianza del error para el valor de j=1.

- **Función de autocorrelación**

Hanke y Reitsh (2006) calculada como cociente entre la función de autocovarianza y la varianza, tal y como ya se vio antes, tendríamos:

$$\rho_j = \frac{\gamma_j}{\gamma_0}$$

$$\rho_j = \frac{\theta_1 \sigma_a^2}{(1 - \theta_1^2) \sigma_a^2} = \frac{\theta_1}{(1 - \theta_1^2)}$$

$$\rho_j = 0 \quad j > 1$$

- **La función de autocorrelación parcial**

Se calcularía siguiendo la siguiente expresión, que no demostramos:

$$\hat{\phi}_{jj} = \frac{\theta_1^j (1 - \theta_1^2)}{1 - \theta_1^{2(j+1)}}$$

Definidas las funciones características de los procesos estocásticos para el caso concreto del MA (1), podemos enunciar las siguientes particularidades de este tipo de proceso:

1. Siempre es estacionario.
2. Para ser invertible, es necesario que $|\theta| < 1$.
3. La ρ_j sólo tiene un punto significativo. El modelo "olvida" la correlación con períodos distintos al inmediatamente anterior y el correlogramas sólo tendrá un punto significativo.
4. La función de autocorrelación parcial no se anula, pero tendrá un comportamiento amortiguado hacia cero.

Modelos AR (1)

Definido el modelo AR (1) como:

$$y_t = \theta_0 + \theta_1 y_{t-1} + e_t$$

Que también puede escribirse como:

$$\theta(L)y_t = \theta_0 + e_t$$

$$\theta(L) = 1 - \theta_1 L$$

A diferencia de los modelos de medias móviles, los autorregresivos no son estacionarios por definición y, para que lo sean, ha de cumplirse que las raíces de la siguiente ecuación sean mayores que uno:

$$\theta(L) = 1 - \theta_1 L = 0$$

Lo que nos permitiría escribirlo como un modelo de medias móviles y, en definitiva, esto supone que los coeficientes $|\theta|$ han de ser menores que 1.

Pasamos a describir las funciones definidas:

- **Esperanza matemática**

$$E(y_t) = E(\theta_1 y_{t-1} + \theta_0 + e_t) =$$

$$\theta_1 E(y_{t-1} + \theta_0 + 0)$$

Donde, como el proceso es estacionario, las esperanzas

$$E(y_t) = E(y_{t-1}) = \dots = E(y_{t-j}) = \mu$$

Y se puede escribir:

$$E(y_t) = \mu = \theta_1\mu + \theta_0$$

$$\mu = \frac{\theta_0}{1 - \theta_1}$$

- Varianza

Para hacer los cálculos con mayor facilidad es conveniente poner el modelo autorregresivo en desviaciones a la media, sin que ello suponga ningún cambio en éste (se puede hacer la prueba escribiendo el modelo en desviaciones y llegando al modelo normal).

El modelo en desviaciones a la media lo definiremos como:

$$\tilde{y}_t = \theta_1\tilde{y}_{t-1} + e_t$$

Dónde:

$$\tilde{y}_{t-j} = y_{t-j} - \mu = y_{tj} - \frac{\theta_0}{1-\theta_1}$$

Para calcular la varianza, escribimos el momento de segundo orden:

$$var(y_t) = \gamma_p = E(\theta_1\tilde{y}_{t-1} + e_t)^2$$

$$E(\theta_1^2\tilde{y}_{t-1}^2 + e_t^2 + 2\theta_1\tilde{y}_{t-1}e_t) =$$

$$\theta_1^2E(\tilde{y}_{t-1})^2 + E(e_t)^2 + 2\theta_1E(\tilde{y}_{t-1}e_t)$$

Dado que el proceso es estacionario:

$$E(\tilde{y}_t)^2 = E(\tilde{y}_{t-1})^2 = \gamma_0$$

$$E(e_t)^2 = \sigma_e^2$$

Como el proceso en desviaciones a la media se puede escribir como un proceso de medias móviles y por lo que hemos visto anteriormente, podríamos escribir:

$$E(e_t \tilde{y}_{t-h}) = \sigma_e^2 \quad / \quad h = 0 \quad \text{y} \quad 0 \quad \text{si} \quad h > 0$$

Por todo lo cual:

$$\gamma_0 = \theta_1^2 \gamma_0 + \sigma_e^2$$

$$\gamma_0 = \frac{\sigma_e^2}{1 - \theta_1^2}$$

- **Función de autocovarianza**

$$\gamma_1 = \text{cov}(\tilde{y}_t \tilde{y}_{t-1}) = E(\tilde{y}_{t-1}(\theta_1 \tilde{y}_{t-1} + e_t)) =$$

$$\theta_1 \gamma_0 = \theta_1 \frac{\sigma_e^2}{1 - \theta_1^2}$$

$$\gamma_2 = \text{cov}(\tilde{y}_t \tilde{y}_{t-2}) = E(\tilde{y}_{t-2}(\theta_1 \tilde{y}_{t-1} + e_t)) =$$

$$\theta_1 \gamma_1 = \theta_1^2 \gamma_0 = \theta_1^2 \frac{\sigma_e^2}{1 - \theta_1^2}$$

Lo que generalizando, se podría escribir como:

$$\gamma_j = \frac{\sigma_e^2}{1 - \theta_1^2} \theta_1^j$$

- **Función de autocorrelación total**

$$\sigma_j = \theta_1^j \quad / \quad j \geq 1$$

- **Función de autocorrelación parcial**

$$\theta_{jj} = \rho_1 \quad / \quad j=1 \quad \text{y} \quad 0 \quad \text{si} \quad j > 1$$

Definidas todas estas funciones, podemos caracterizar el proceso autorregresivo del siguiente modo:

1. Siempre es invertible (está directamente invertido).
2. Para ser estacionario, ha de cumplirse que $|\theta| < 1$.
3. La función de autocorrelación total no se anula, pero se va amortiguando hacia cero.
4. La función de autocorrelación parcial se anula para retardos superiores a uno.

Identificación del modelo

Aunque podríamos seguir definiendo las características de otros procesos ARIMA de órdenes mayores, no tiene mayor interés una vez se ha entendido el procedimiento y sí interesa precisar la forma que adoptarían los correlogramas de estas funciones porque, fruto de su comparación con los que obtendremos con nuestras series de interés, podremos asociar a nuestra serie temporal de estudio un proceso ARIMA que identifique su proceso generador de datos, tanto a pasado como a futuro.

Los correlogramas o funciones de autocorrelación total y parcial están en la figura 1, 2, 3 también disponibles en el libro de (Hanke y Reitsh, 2006).

De forma muy poco académica, el proceso de identificación consistirá en calcular las funciones de autocorrelación total y parcial de nuestra serie (una vez estamos seguros de que ésta cumple las condiciones que definen un proceso estocástico estacionario) y comparar sus correlogramas con los correspondientes a los modelos teóricos AR (p), MA (q) o ARMA (p, q).

En principio, si el proceso está bien identificado, procederemos a su estimación y, si analizamos los correlogramas de los residuos obtenidos en la estimación, serán "ruido blanco". Si esto no es así, habrá que realizar una nueva estimación incorporando la estructura más parecida al modelo teórico que podamos intuir con la comparación con los modelos teóricos.

Para saber cuándo estamos ante un "ruido blanco", se pueden hacer las siguientes comprobaciones:

- **Media nula**

Universidad Autónoma de Madrid (2015) puede observarse en el gráfico de residuos si el error se mueve en torno al valor cero o bien calcularse el cociente entre la media y la varianza muestral de los residuos. Si ese ratio es inferior a 2, podemos concluir (con un $\varepsilon=0,05$) que la media no es significativamente distinta de cero.

- **Varianza constante**

Universidad Autónoma de Madrid (2015) Observando el gráfico de los residuos puede analizarse la constancia de la varianza del error. En caso de heterocedasticidad y es recomendable una transformación logarítmica en la serie original.

- **Incorrelación**

Deben observarse los coeficientes de autocorrelación muestral de los residuos y comprobar que ninguno de ellos supera el valor de las bandas de significatividad al 5% ($\pm 1,96 (1/T^{1/2})$). El valor $T^{1/2}$ es una aproximación de la varianza asintótica pero resulta sólo adecuada para valores grandes de "j". El estudio de las funciones de autocorrelación muestral y autocorrelación parcial muestral de los residuos, pueden servirnos fácilmente para el replanteamiento del modelo inicial.

El contraste de la "Q" de Box-Pierce analiza la hipótesis nula de que:

$$H_0: \rho_1(a) = \rho_2(a) = \rho_3(a) = \dots = \rho_M(a) = 0$$

Suponiendo que la expresión:

$$Q = T \sum_{j=1}^M r_j^2(\hat{a})_t$$

O la alternativa propuesta por Ljung-Box:

$$Q^* = T(T+2) \sum_{j=1}^M (T-j)^{-1} r_j^2(\hat{e})_t$$

Se distribuye como una chi-cuadrado con M-k grados de libertad.

- **Procesos ARMA**

- Los procesos AR no pueden representar series de memoria muy corta, donde el valor actual de la serie solo está correlacionado con un número pequeño de valores anteriores de la serie.
- Los procesos MA son función de un número finito de innovaciones anteriores, no de todas.
- La clase de modelos ARMA es amplia y flexible, ya que combina las estructuras AR y MA. Es útil para representar una gran variedad de series utilizando pocos parámetros.

Universidad Carlos III de Madrid (01 de octubre de 2008) Predicción con modelos ARIMA [Departamento de Estadística]. Recuperado de http://portal.uc3m.es/portal/page/portal/dpto_estadistica/personal/isabel_molina_peralta

Uriel (1985) la presentación de procesos AR y MA da lugar a los procesos mixtos ARMA. La formulación general de un proceso ARMA, ARMA (p, q), es:

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$$

ARMA (1,1)

Un proceso ARMA (1,1) (se excluye la constante por simplicidad):

$$Y_i = \varphi_1 Y_{i-1} + e_t - \theta_1 e_{t-1}$$

- Modelo de medias móviles

Un modelo de los denominados de medias móviles es aquel que explica 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.

Uriel (1985) estos modelos se denotan normalmente con las siglas MA, seguidos, como en el caso de los modelos autorregresivos, del orden entre paréntesis. Así, un modelo con q términos de error MA (q) respondería a la siguiente expresión:

$$Y_t = c + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}$$

Que de nuevo puede abreviarse utilizando el polinomio de retardos (como en caso de los modelos AR):

$$Y_t = \theta_q(L)e_t + \mu$$

Al igual que en el caso de los modelos autorregresivos, el orden de los modelos de medias móviles suele ser bajo MA(1), MA(2) o corresponderse con la periodicidad de los datos analizados MA(4), para series trimestrales, o MA(12) para series mensuales.

Interpretación de un modelo de medias móviles

Universidad Autónoma de Madrid (2015) así como un modelo autorregresivo es intuitivamente sencillo de comprender, la formulación de un modelo de medias móviles resulta sorprendente para el no iniciado. ¿Qué significa que una variable aleatoria se explique en función de los errores cometidos en períodos precedentes?, ¿De dónde proceden esos errores?, ¿Cuál es la justificación de un modelo de este tipo?

En realidad, un modelo de medias móviles puede obtenerse a partir de un modelo autorregresivo sin más que realizar sucesivas sustituciones.

- Modelos MA(1)

Espaza y Cancelo (1993) una vez tenemos definidas las ventajas de contar con un proceso estocástico estacionario lineal y discreto, y que podemos calcular las funciones de autocovarianza y autocorrelación, puede resultar interesante ver que valores toman éstas en aquellos casos sencillos que luego nos permitan comprobar si series económicas generales pueden tener un comportamiento similar, simplemente acudiendo a la comparación de sus correlogramas (de la función de autocorrelación total y parcial).

El primer caso a analizar será el modelo de medias móviles de orden uno, que se define como:

$$Y_t = \mu + e_t + \theta_1 e_{t-1}$$

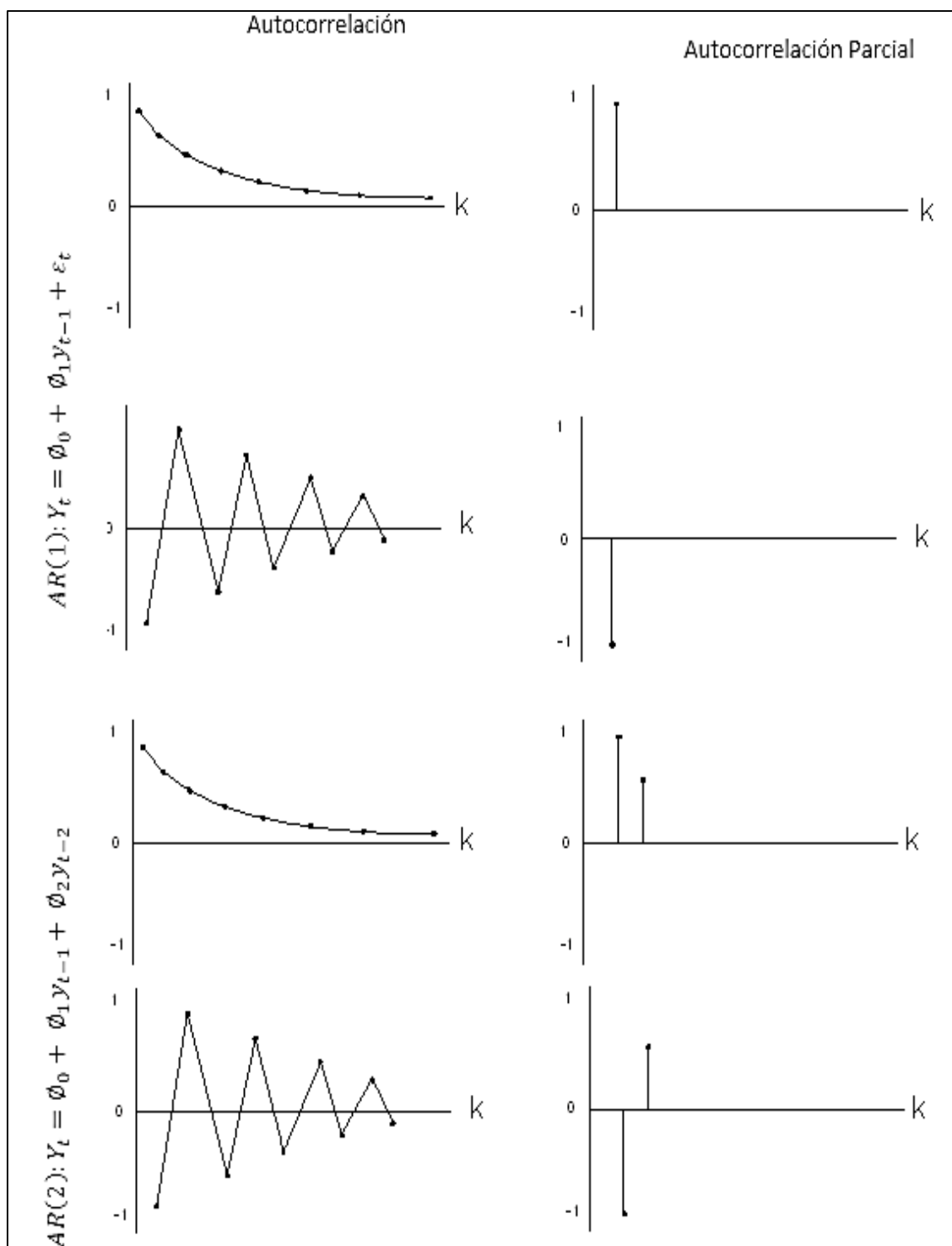


Figura 1: Coeficiente de autocorrelación y de autocorrelación parcial de los modelos AR(1) y AR(2).

FUENTE: (Hanke y Reitsh, 2006). Pronósticos en los negocios (p. 433).

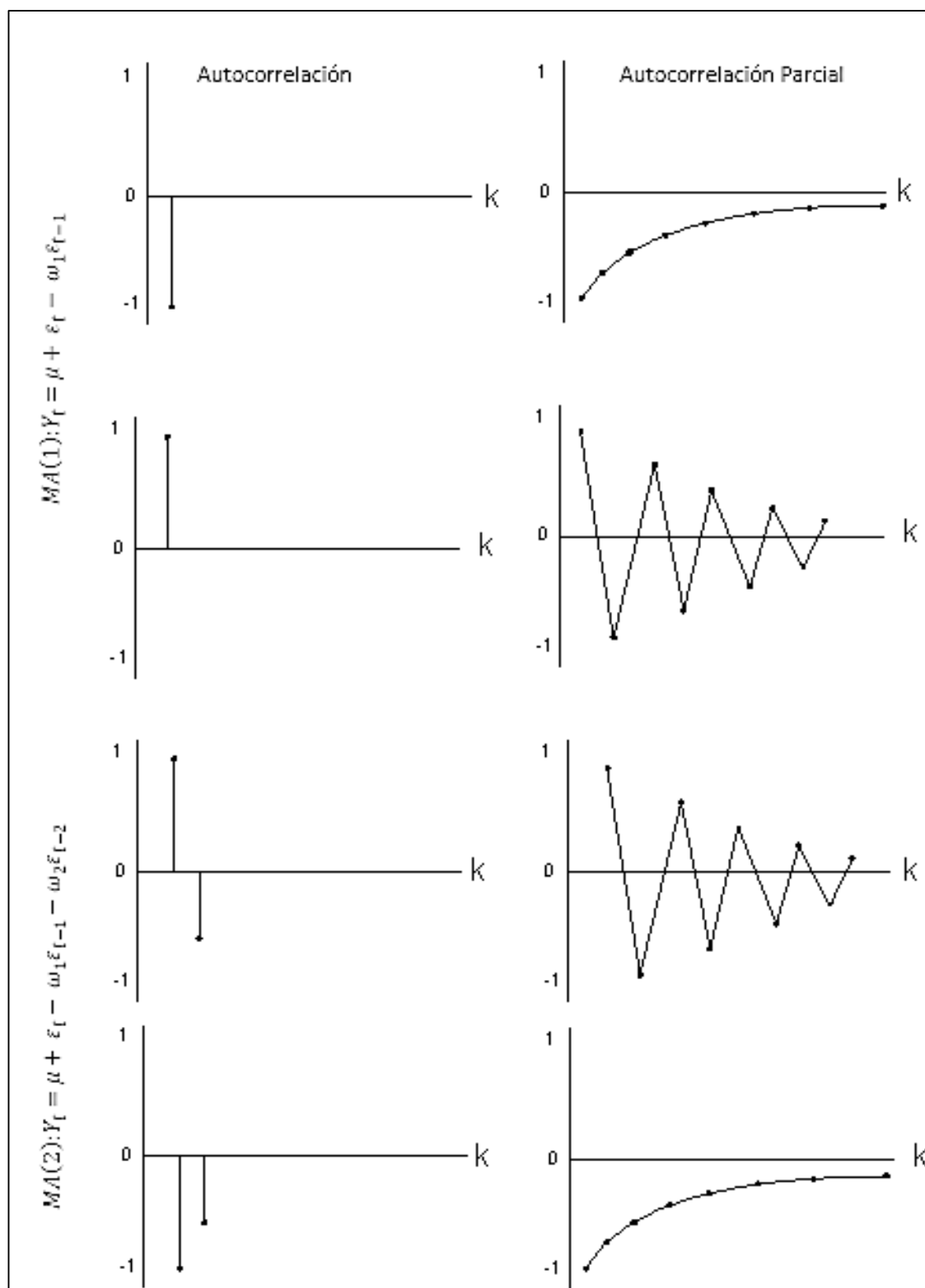


Figura 2: Coeficiente de autocorrelación y de autocorrelación parcial de los modelos MA(1) y MA(2).

FUENTE: (Hanke y Reitsh, 2006). Pronósticos en los negocios (p. 434).

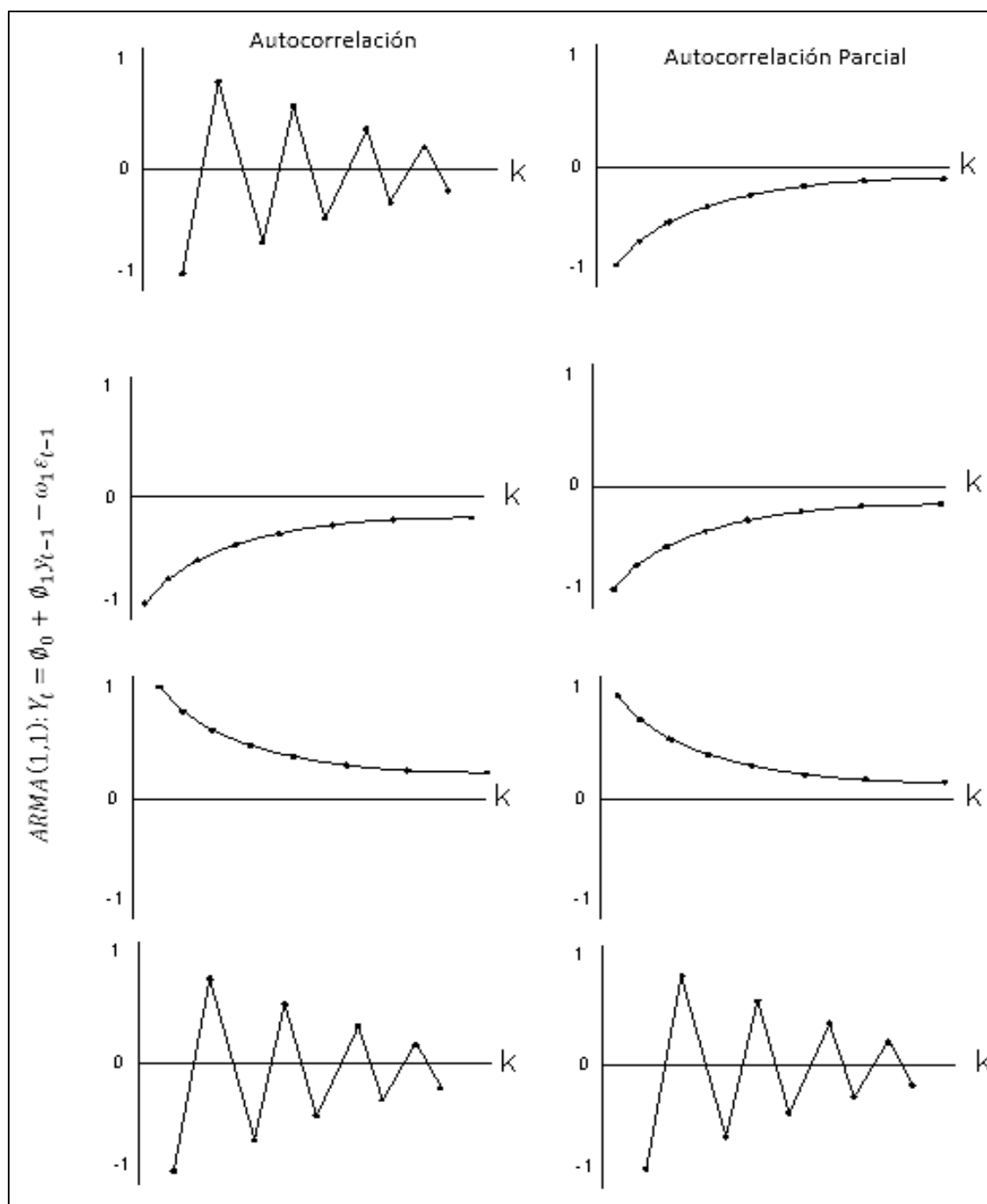


Figura 3: Coeficientes de autocorrelación y autocorrelación parcial de un modelo ARMA(1, 1).

FUENTE: (Hanke y Reitsh, 2006). Pronósticos en los negocios (p. 435).

Modelo Box Jenkins

Enfoque del método

El método original utiliza un enfoque de modelado iterativo en tres etapas, usando datos de un horno de gas. Estos datos son conocidos como datos de Box-Jenkins del horno de gas para la evaluación comparativa de modelos de predicción.

Las tres etapas del modelado iterativo son las siguientes:

1. **Identificación y selección del modelo:** asegurarse de que las variables son estacionarias, la identificación de la estacionalidad de la serie dependiente (diferenciando en temporada si es necesario), y el uso de los gráficos de las funciones de autocorrelación y de autocorrelación parcial de la serie de tiempo dependiente para decidir cuál componente (si es el caso) se debe utilizar en el modelo, el promedio autorregresivo o un promedio móvil.
2. **Estimación de parámetros usando algoritmos de cálculo para tener coeficientes que mejor se ajustan al modelo ARIMA seleccionado.** Los métodos más comunes usan estimación de máxima verosimilitud o mínimos cuadrados no lineales.
3. **Comprobar el modelo mediante el ensayo,** si el modelo estimado se ajusta a las especificaciones de un proceso univariado estacionario. En particular, los residuos deben ser independientes el uno del otro y la media y la varianza constantes en el tiempo. (Graficando la media y la varianza de los residuos a través del tiempo y la realización de una prueba de Ljung-Box o el trazado de autocorrelación y autocorrelación parcial de los residuos son útiles para identificar los errores de

especificación.) Si la estimación es suficiente, tenemos que volver al paso uno y el intento de construir un modelo mejor.

Identificación del modelo Box Jenkins

Estacionariedad y estacionalidad

El primer paso en el desarrollo de un modelo de Box-Jenkins es determinar si la serie de tiempo es estacionaria y si hay alguna estacionalidad significativa que necesite ser modelada.

Detección estacionariedad

La estacionariedad puede evaluarse a partir de una secuencia de largo plazo. La secuencia ejecutada debe mostrar ubicación y escala constante. También se puede detectar a partir de una secuencia de autocorrelación. En concreto, la no estacionariedad se indica a menudo por un gráfico de autocorrelación con una caída muy lenta.

Detección estacionalidad

La Estacionalidad (o periodicidad) generalmente se puede evaluar a partir de un diagrama de autocorrelación, una sub serie de una temporada, o una trama espectral.

Diferenciación para lograr estacionariedad

Box y Jenkins recomiendan el enfoque de diferenciación para lograr estacionariedad. Sin embargo, ajustando una curva y restando los valores ajustados de los datos originales también se puede utilizar en el contexto de los modelos de Box-Jenkins.

Diferenciación estacional

En la fase de identificación del modelo, el objetivo es detectar la estacionalidad, si existe, y para identificar el orden de los términos autorregresivos y de media móvil estacional de temporada. Para muchas series, el período es conocido y un solo término estacional es suficiente. Por ejemplo, para los datos mensuales que uno suele incluir un término de temporada AR 12 o un término de temporada MA 12. Para los modelos de Box-Jenkins, uno no elimina explícitamente estacionalidad antes de ajustar el modelo. En su lugar, se incluye la orden de los términos estacionales en la especificación del modelo a la ARIMA software de estimación. Sin embargo, puede ser útil aplicar una diferencia estacional a los datos y regenerar la autocorrelación y autocorrelación parcial parcelas. Esto puede ayudar en la identificación del modelo del componente no estacional del modelo. En algunos casos, la diferenciación estacional puede eliminar la mayor parte o la totalidad del efecto de la estacionalidad.

Identificar P Y Q

Una vez que se han abordado estacionalidad y temporalidad, el siguiente paso es identificar el orden (es decir, la p y q) de los términos autorregresivo y de media móvil. Diferentes autores tienen diferentes enfoques para la identificación de p y q . Brockwell y Davis (1991, p. 273) afirman "nuestro principal criterio para la selección del modelo de [entre ARMA (p , q) los modelos] será la AICC", es decir, el criterio de información Akaike con corrección.

Otros autores utilizan el argumento de autocorrelación y autocorrelación parcial de la parcela.

Autocorrelación y autocorrelación parcial

La muestra autocorrelación trama y la muestra gráfica de autocorrelación parcial se comparan con el comportamiento teórico de estas parcelas cuando el pedido es conocido.

Específicamente, para un (1) AR proceso, la función de autocorrelación de la muestra debe tener una apariencia de forma exponencial decreciente. Sin embargo, los procesos AR de orden superior son a menudo una mezcla de forma exponencial decreciente y amortiguados componentes sinusoidales.

Para procesos autorregresivos de orden superior, la autocorrelación de la muestra tiene que ser complementado con una parcela de autocorrelación parcial. La autocorrelación parcial de un AR (p) el proceso se convierte en cero al retardo $p + 1$, y mayor, por lo que examinar la función de autocorrelación parcial de la muestra para ver si hay evidencia de una desviación de cero. Esto generalmente se determina mediante la colocación de un 95% intervalo de confianza de la muestra gráfica de autocorrelación parcial (la mayoría de los programas de software que generan muestras parcelas autocorrelación también trazar este intervalo de confianza). Si el programa de software no genera la banda de confianza, que es de aproximadamente $\pm 2/\sqrt{N}$, con N que denota el tamaño de la muestra.

La función de autocorrelación de un MA (q) el proceso se convierte en cero al retardo $q + 1$, y mayor, por lo que examinar la función de ejemplo de autocorrelación para ver donde esencialmente se convierte en cero. Hacemos esto mediante la colocación de un intervalo de confianza del 95% para la función de ejemplo de autocorrelación en la parcela de muestreo de autocorrelación. La mayoría del software que genera el diagrama de autocorrelación también puede generar este intervalo de confianza.

La función de autocorrelación parcial de la muestra en general, no es útil para identificar el orden del proceso de media móvil.

En la práctica, la autocorrelación de la muestra y las funciones de autocorrelación parcial son variables aleatorias y no dar la misma imagen que las funciones teóricas. Esto hace que el modelo de identificación más difícil. En particular, los modelos mixtos pueden ser particularmente difíciles de identificar. Aunque la experiencia es de gran ayuda, el desarrollo de buenos modelos que utilizan estas parcelas se implica mucho ensayo y error.

Modelo de estimación de box jenkins

La estimación de los parámetros de los modelos de Box-Jenkins es un problema de estimación no lineal bastante complicado. Por esta razón, la estimación de parámetros debe dejarse a un programa de software de alta calidad que se ajuste a los modelos de Box-Jenkins. Afortunadamente, muchos programas de software estadístico ahora encajan modelos Box-Jenkins.

Los principales enfoques de los modelos de Box-Jenkins montaje son mínimos cuadrados no lineales y la estimación de máxima verosimilitud. Estimación de máxima verosimilitud es generalmente la técnica preferida. Las ecuaciones de verosimilitud para el modelo completo de Box-Jenkins son complicados y no se incluyen aquí. Ver (Brockwell y Davis, 1991) para los detalles matemáticos. El diagnóstico de modelos Box-Jenkins Supuestos para un proceso univariado estable

Diagnósticos modelo para los modelos de Box-Jenkins es similar al modelo de validación de mínimos cuadrados no lineales de montaje.

Es decir, el término de error u_t se supone que sigue los supuestos para un proceso univariado estacionario. Los residuos deben ser ruido blanco (o independientes cuando sus distribuciones son normales) dibujos a partir de una distribución fija con una media constante y varianza. Si el modelo de Box-Jenkins es un buen modelo para los datos, los residuos deben satisfacer estos supuestos.

Si estos supuestos no se cumplen, hay que adaptarse a un modelo más apropiado. Es decir, volver a la etapa de identificación del modelo y tratar de desarrollar un modelo mejor. Esperemos que el análisis de los residuos puede dar algunas pistas en cuanto a un modelo más apropiado.

Una forma de evaluar si los residuos del modelo de Box-Jenkins siguen las hipótesis es la de generar gráficos estadísticos de los residuales. También se podría mirar el valor del estadístico Box-Ljung. Wikipedia (29 de octubre de 2016). Metodología Box-Jenkins [Wikipedia]. Recuperado de https://es.wikipedia.org/wiki/Metodolog%C3%ADa_de_Box-Jenkins

Criterio de información akaike (AIC)

El criterio de información de Akaike es una medida de la bondad de ajuste de un modelo estadístico. Se puede decir que describe la relación entre el sesgo y varianza en la construcción del modelo, acerca de la exactitud y complejidad del modelo.

El AIC no es una prueba del modelo en el sentido de prueba de hipótesis. Más bien, proporciona un medio para la comparación entre los modelos de una herramienta para la selección del modelo. Dado un conjunto de datos, varios modelos candidatos pueden ser clasificados de acuerdo a su AIC, con el modelo que tiene el mínimo AIC es la mejor.

A partir de los valores de la AIC también se puede inferir que, por ejemplo, los dos primeros modelos están más o menos empatados y el resto son mucho peores.

1. En general, el AIC se define como:

$$AIC = 2k - 2 * \ln(L)$$

Donde:

- **k** es el número de parámetros del modelo
 - **ln(L)** es la función de log-verosimilitud para el modelo estadístico
2. Para los conjuntos de datos más pequeños, la AIC_c se aplica la corrección de segundo orden:

$$AIC_c = AIC + \frac{2k(k+1)}{N-k-1} = \frac{2 * N * k}{N-k-1} - 2 * \ln(L)$$

Donde:

- **N** es el tamaño de la muestra de datos.
- **k** es el número de parámetros del modelo.

CAPÍTULO III

MATERIALES Y MÉTODOS

El presente trabajo de investigación se realizó en el Hospital Regional Manuel Núñez Butrón (MNB) de la Provincia y Departamento de Puno a 3819 m.s.n.m.

3.1. POBLACIÓN

La población en estudio está conformada por el número total de pacientes registrados mensualmente en el libro de informes de los diferentes consultorios externos en el Hospital Regional Manuel Núñez Butrón – Puno.

3.2. MUESTRA

La muestra está conformada por el número de pacientes registrados mensualmente, elegidas a criterio por la disponibilidad de los datos, el periodo de tiempo comprendido es desde el año 2008 hasta el año 2017, en datos mensuales $n=120$.

3.3. UNIDAD MUESTRAL

Número de atenciones en los consultorios externos como son: MEDICINA INTERNA, GINECOLOGÍA Y OBSTETRICIA, CIRUGÍA, MEDICINA FÍSICA Y REHABILITACIÓN, PEDIATRÍA, OFTALMOLOGÍA

3.4. METODO DE RECOLECCION DE DATOS

Con la finalidad de asegurar la efectividad y originalidad de los datos estos fueron recopilados directamente de la base de datos de la oficina de Estadística del Hospital Regional Manuel Núñez Butrón – Puno 2008 – 2017.

3.5. METODOLOGIA

El presente trabajo de investigación es analítico – aplicativo; este método es de suma importancia, ya que está presente en todas las fases de los modelos

3.6. METODO DE ANALISIS DE DATOS

El presente trabajo de investigación se utilizó la teoría de WIENER-KOLMOGOROV, más conocido como el enfoque de Box-Jenkins en las series de tiempo.

Los pasos a seguir en la obtención del modelo univariante por el método Box-Jenkins fueron:

- a. Representación gráfica de las series
- b. Calculo de la función autocorrelación (F.A.C.) y función de autocorrelación parcial (F.A.C.P.).
- c. Proceso de identificación.
- d. Estimación de parámetros.
- e. Proceso de verificación.
- f. Proceso de predicción.

3.6.1. Método de box – jenkins (teoría de wiener – kolmogorov).

La metodología de Box-Jenkins sigue un proceso que consta de cuatro fases, las cuales son:

1. Identificación.- Se trata de elegir uno o varios modelos ARIMA como posibles candidatos para explicar el comportamiento de la serie.

2. Estimación.- Se realiza la estimación de los parámetros de los modelos seleccionados.
3. Verificación o Validación.- Se comprueba la adecuación del modelo estimado y se verifica que cumpla las propiedades respectivas.
4. Pronóstico y Predicción.- Si el modelo elegido es satisfactorio se realizan las predicciones de la variable.

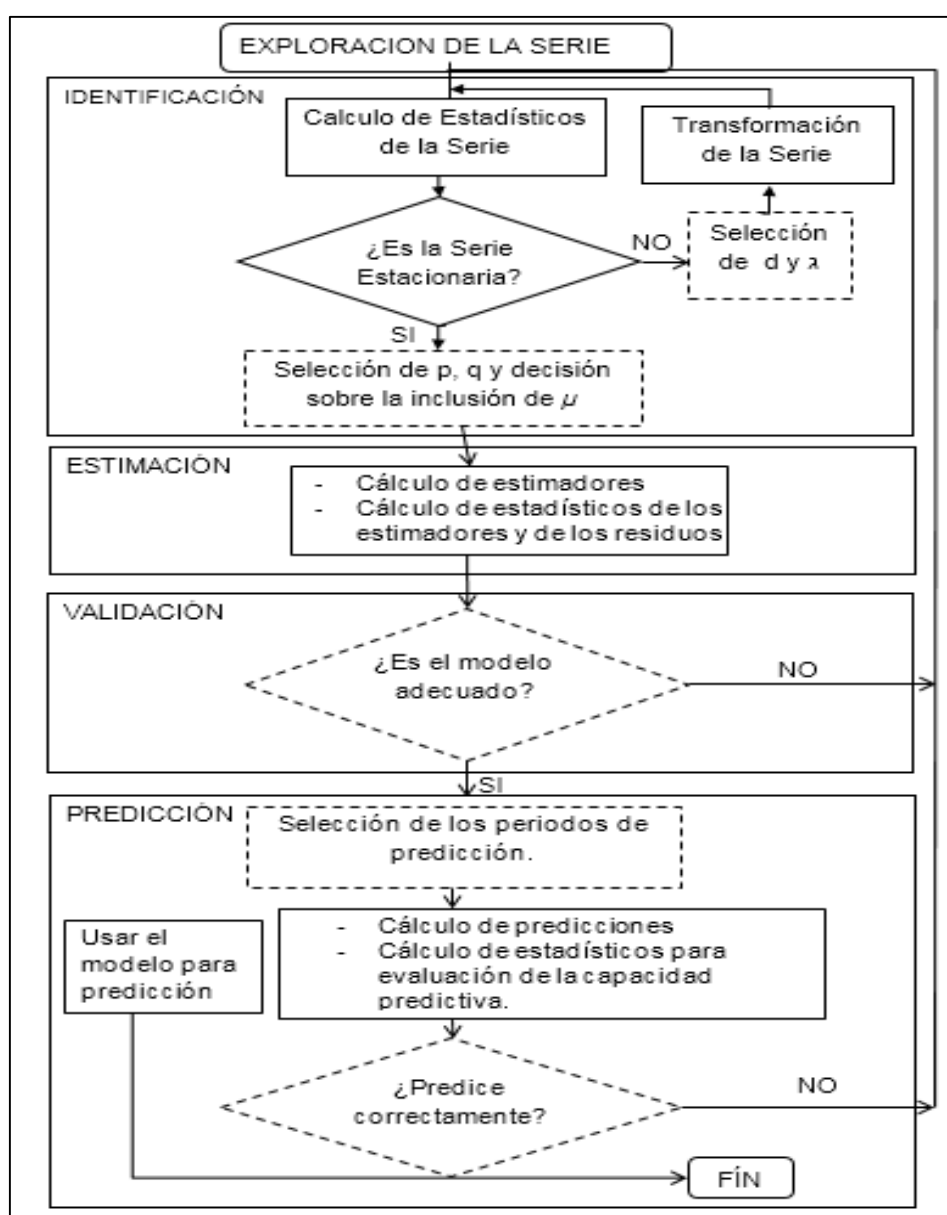


Figura 4: Metodología del enfoque Box – Jenkins

FUENTE: Ezequiel, (1985). Análisis de series temporales: Modelos ARIMA (p. 72).

Fase 1: Identificación de modelos

Anderson (1985) primero, tenemos en cuenta que el modelo ARIMA puede ser considerado para un caso donde se presenta una adecuada representación de los procesos generados de una serie de tiempo Y_1, \dots, Y_N . La media de la serie, la varianza, la F.A.C. y la F.A.C.P. son calculados, y la inspección de r_k y ϕ_{kk} debería indicar el (los) modelo(s) a ser dispares (entretenido)

El orden de integración (o grado de diferenciación), denotado por $I(d)$, se refiere al número de veces que una serie sea diferenciada para optar una serie estacionaria. El orden de integración define el parámetro “d” del modelo ARIMA

Se trata de la determinación de estacionariedad de la serie (d y λ) y a continuación el número de parámetros (p) y la media móvil (q), es decir si el modelo de la media varía a través del tiempo, se trata de series no estacionarias se aplican las transformaciones adecuadas con la finalidad de convertir en estacionarias e invertibles, especificando el grado de diferenciación y el algoritmo de Box-Jenkins haciendo uso de los siguientes.

- Representación gráfica de la serie; se visualiza fluctuaciones respecto a la media para confirmar la estacionariedad de la serie.
- Estimación de la función de autocorrelación (F.A.C.) y la función de autocorrelación parcial (F.A.C.P.); se demuestra la significancia de los r_k y ϕ_{kk} confirmar que ninguno de los parámetros estimados sea superior a 1 ni menor a $(q-1)$.
- Calcular las raíces de la ecuación característica; en el proceso de identificación se comprueba la estacionariedad de la serie, solamente si las raíces caen dentro del

circulo unitario, es conveniente realizar esta inspección. Para esto se utiliza la prueba de Dickey Fuller o de raíces unitarias en la cual la hipótesis a cotejar es:

H_0 : Hay raíz unitaria (proceso no estacionario).

H_1 : No hay raíz unitaria (proceso estacionario).

El estadístico de prueba es: $T_u = \frac{a}{\hat{s}(\hat{a}_0)}$

Donde:

\hat{a}_0 : Estimación mínimo cuadrado de a . La regla de decisión a considerar será:

Rechazar H_0 si $T_u > T_t$ y por tanto el proceso es estacionario. El estadístico T_u sigue una distribución T de Students que se tabulo por Dickey y Fuller

Fase 2: Estimación de parámetros

Una vez seleccionado el modelo tentativo, habiendo identificado los valores apropiados de p , d y q se procede a estimar los valores de parámetros de los términos autorregresivos y media móvil incluidos en el modelo, a través del uso de los mínimos cuadrados no lineales que minimizan la suma de los cuadrados de los residuos.

Fase 3: Validación o verificación

La finalidad de esta fase consiste en analizar la adecuación entre el modelo y los datos, o dicho de otra forma en que medida se cumple lo siguiente:

- a) Los residuos del modelo estimado se aproximan al comportamiento de un ruido blanco.
- b) El modelo estimado es estacionario e invertible.

- c) Los coeficientes son estadísticamente significativos.
- d) El grado de ajuste es elevado en comparación al de otros modelos alternativos.
Una de las formas para detectar violaciones a los supuestos es a través del análisis de los residuales, que presentan la parte de las observaciones que no es explicada por el modelo (uso del estadístico de Box-Pierce).

Conviene señalar que es primordial que se cumpla el requisito a). Pues en caso contrario el modelo debe ser rechazado.

Fase 4: Predicción o pronóstico

- Una vez que se encontró un modelo adecuado se puede utilizar predicciones para uno o varios periodos de futuro.
- Al haber más datos disponibles, se puede utilizar el mismo modelo para las predicciones, selección de otro periodo de origen.
- Si la serie parece cambiar a través del tiempo como pudiera ser necesario de calcular los parámetros o incluso desarrollar un modelo nuevo por completo.

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

4.1. ANÁLISIS DE LA SERIE MENSUAL DEL NÚMERO DE PACIENTES ATENDIDOS EN LOS CONSULTORIOS EXTERNOS DEL HOSPITAL REGIONAL MANUEL NÚÑEZ BUTRÓN

Se presenta en la tabla 1 el consolidado de la serie mensual del número de pacientes atendidos en el Hospital Regional Manuel Núñez Butrón de Puno, correspondientes a los periodos del año 2008 hasta el año 2017, obtenidos de los archivos de la oficina de Estadística del Hospital Regional Manuel Núñez Butrón de Puno.

Tabla 1: Serie del número de pacientes atendidos en consultorios externos, periodo 2008-2017.

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
ENERO	1319	1423	1298	1632	1101	1536	1309	1286	1935	2831
FEBRERO	1006	1175	1221	1463	760	1252	1098	1343	1519	2229
MARZO	1370	1342	1545	1519	1499	1240	1444	1646	1516	2435
ABRIL	1285	1287	1456	1395	1170	1399	1174	1738	1239	2335
MAYO	1062	854	1069	1362	910	1214	493	1644	1166	1905
JUNIO	701	454	935	1473	921	480	427	1442	1675	1545
JULIO	762	369	1432	1944	1244	281	458	2406	1707	781
AGOSTO	1080	602	1559	2463	1455	705	500	2619	2605	1080
SETIEMBRE	855	654	1606	2506	865	844	464	2748	2537	2553
OCTUBRE	677	770	1450	2216	445	908	631	2269	2439	2291
NOVIEMBRE	1397	780	1069	1311	1637	1157	403	1736	893	1706
DICIEMBRE	967	754	1112	1493	1067	867	640	1584	1470	912

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

En la tabla 1 de los datos de pacientes atendidos, se puede observar que el mayor número de pacientes atendidos es 2748, correspondientes al mes de Setiembre del año 2015, con el menor número de pacientes atendidos 281 en el mes de Julio del año 2013, encontrándose los promedios de cada mes de Enero a Diciembre: 1567, 1307, 1556, 1448, 1168, 1005, 1138, 1467, 1563, 1410, 1209, 1087, promedios anuales del año 2008 al 2017: 1040, 872, 1313, 1731, 1090, 990, 753, 1872, 1725, 1884 a continuación se presenta el análisis de la serie del número de pacientes atendidos.

4.1.1. Identificación del modelo

Siguiendo la metodología Box-Jenkins para la modelización de series invariantes se dispone de 120 datos mensuales de la serie histórica del número de pacientes atendidos en el Hospital Regional Manuel Núñez Butrón correspondientes al periodo 2008-2017.

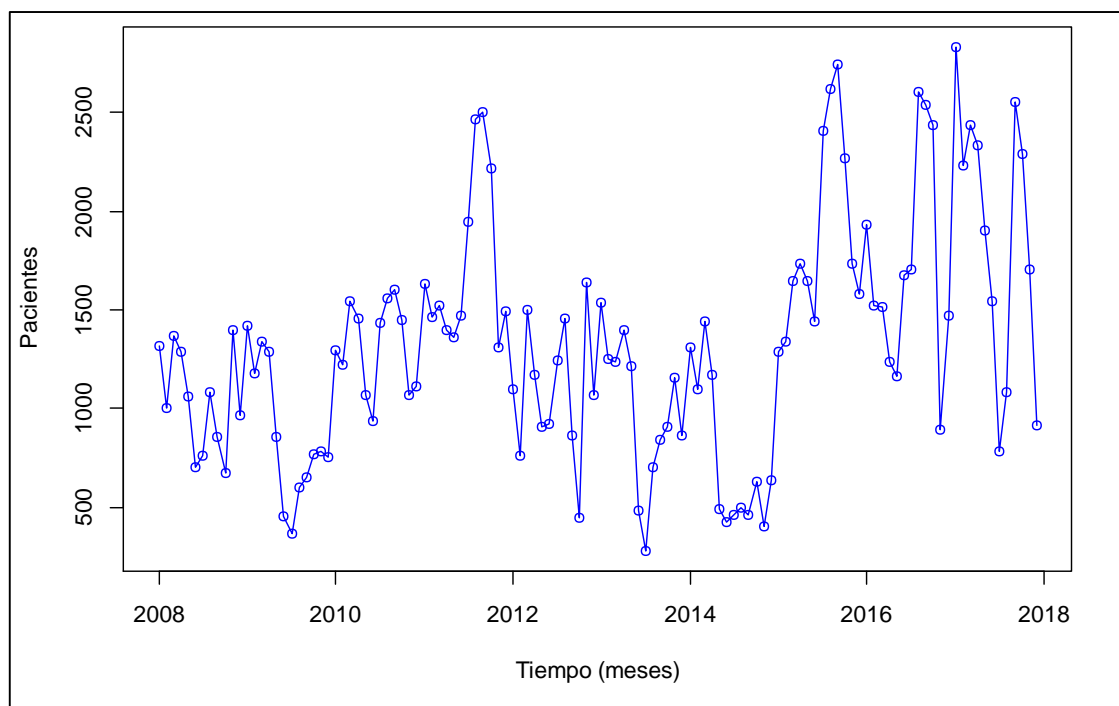


Figura 5: Evolución del número de pacientes atendidos, periodo 2008-2017.

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

En la Figura 5, se observa la evolución de la serie del número de pacientes atendidos, en el periodo 2008-2017 obtenido con el Software estadístico R, ha sido obtenido del software libre. Los modelos Univariantes para describir y predecir la serie histórica del número de atenciones del Hospital Regional Manuel Núñez Butrón – Puno, 2008 – 2017, de la Figura 5, podemos visualizar una variabilidad de sus componentes que muestra muchos picos, muchos de los cuales parecen estar espaciados de manera distante, estos picos indican que nuestra serie tiene mucha variabilidad por lo cual debemos de diferenciarla.

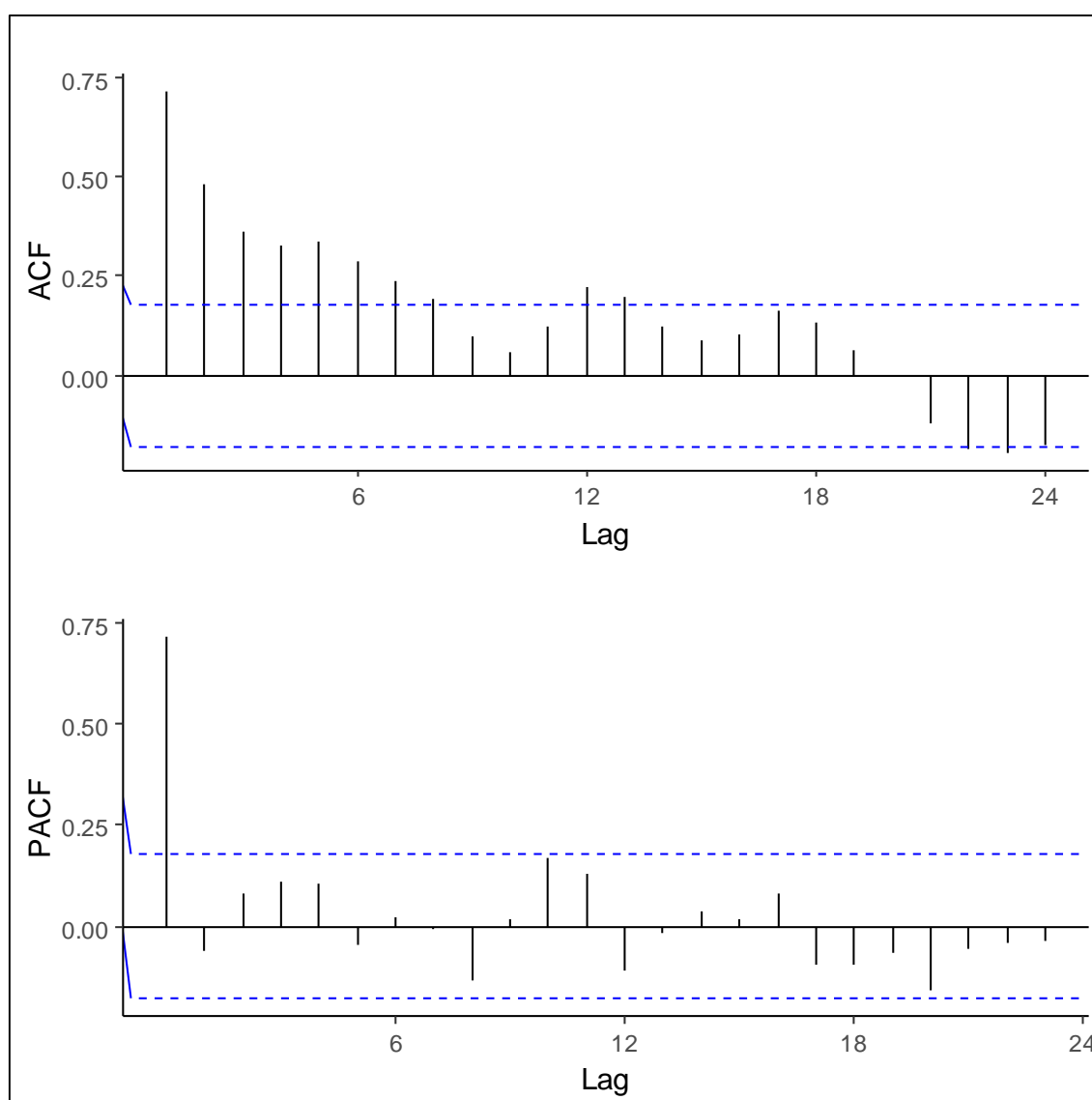


Figura 6: Correlograma FAC y FACP de la serie del número de pacientes atendidos, periodo 2008-2017.

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

En la Figura 6, se obtienen las funciones de autocorrelación y de autocorrelación parcial obtenidas.

También se observa que las funciones de autocorrelación y autocorrelación parcial estimadas también validan los periodos estacionales por que los coeficientes de autocorrelación para retardos múltiplos del periodo estacional de la serie son significativamente distintos de cero.

Prueba analítica de Estacionariedad Dickey-Fuller para la serie del número de pacientes atendidos.

H_0 : $p > 0.05$ La serie es no estacionaria (Tiene raíz unitaria)

H_a : $p < 0.05$ La serie es estacionaria (No tiene raíz unitaria)

Se aplicó el test de Dickey-Fuller y se obtuvo el valor de -2.9182 con un p-valor = 0.1952 que es mayor a nuestro nivel de significancia de 0.05, por ende se acepta la hipótesis nula y se rechaza la hipótesis alterna.

Se concluye que la serie del número de pacientes atendidos no es estacionaria.

Transformación de la serie

Para que la serie del número del número de pacientes atendidos en el Hospital Regional Manuel Núñez Butrón - Puno, 2008-2017, sea estacionaria en media se debe de transformar con la primera diferencia regular a la serie del número de pacientes atendidos y tenemos como resultado lo siguiente:

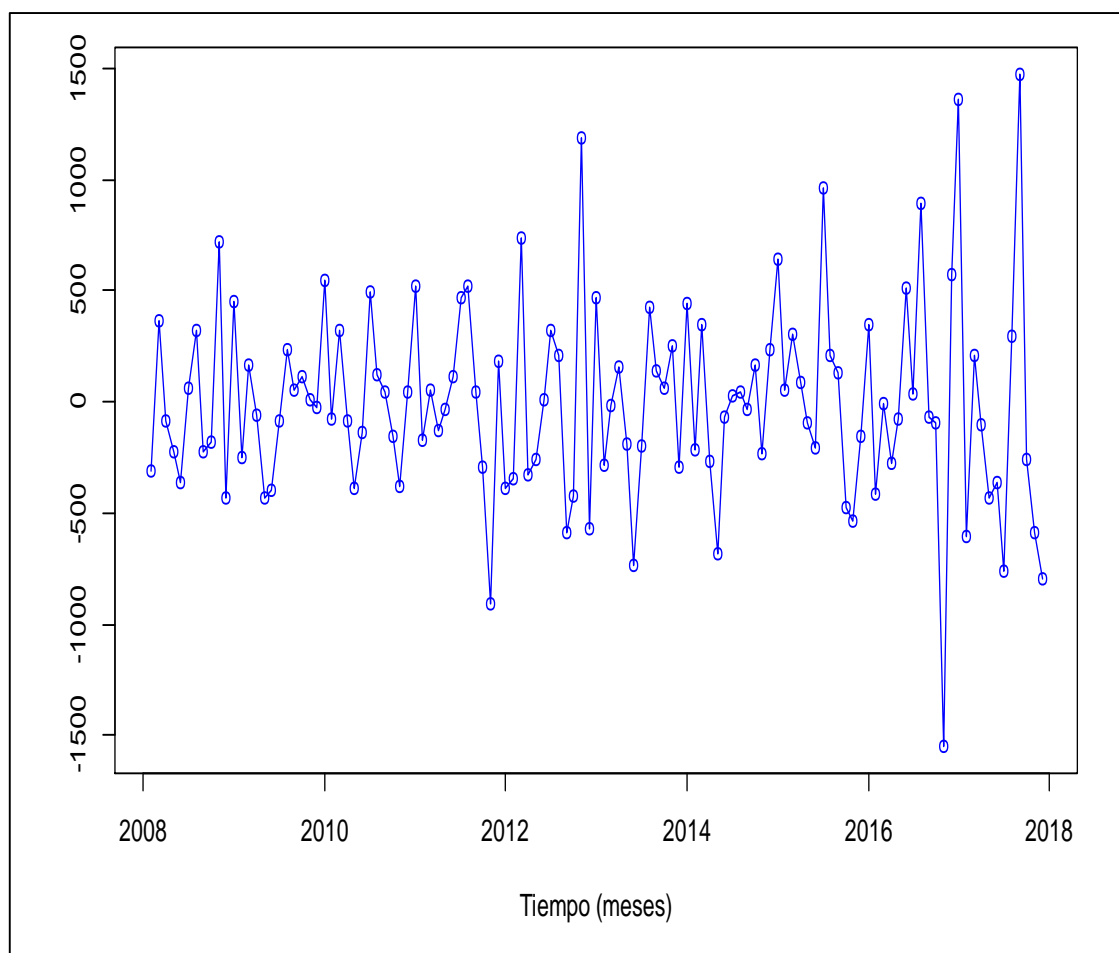


Figura 7: Primera diferencia regular de la serie del número de pacientes atendidos, periodo 2008-2017.

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

En la Figura 7 se puede ver que los datos presentan oscilaciones con esto se presume que la serie es estacionaria.

Se presenta la primera diferencia de la serie histórica no estacional, y tiene la forma de un ruido blanco por lo que dicha serie histórica puede decir que es estacionaria, Después de realizar la primera diferencia regular se pierde un dato, a continuación podemos observar la función de autocorrelación y función de autocorrelación parcial de nuestra serie transformada.

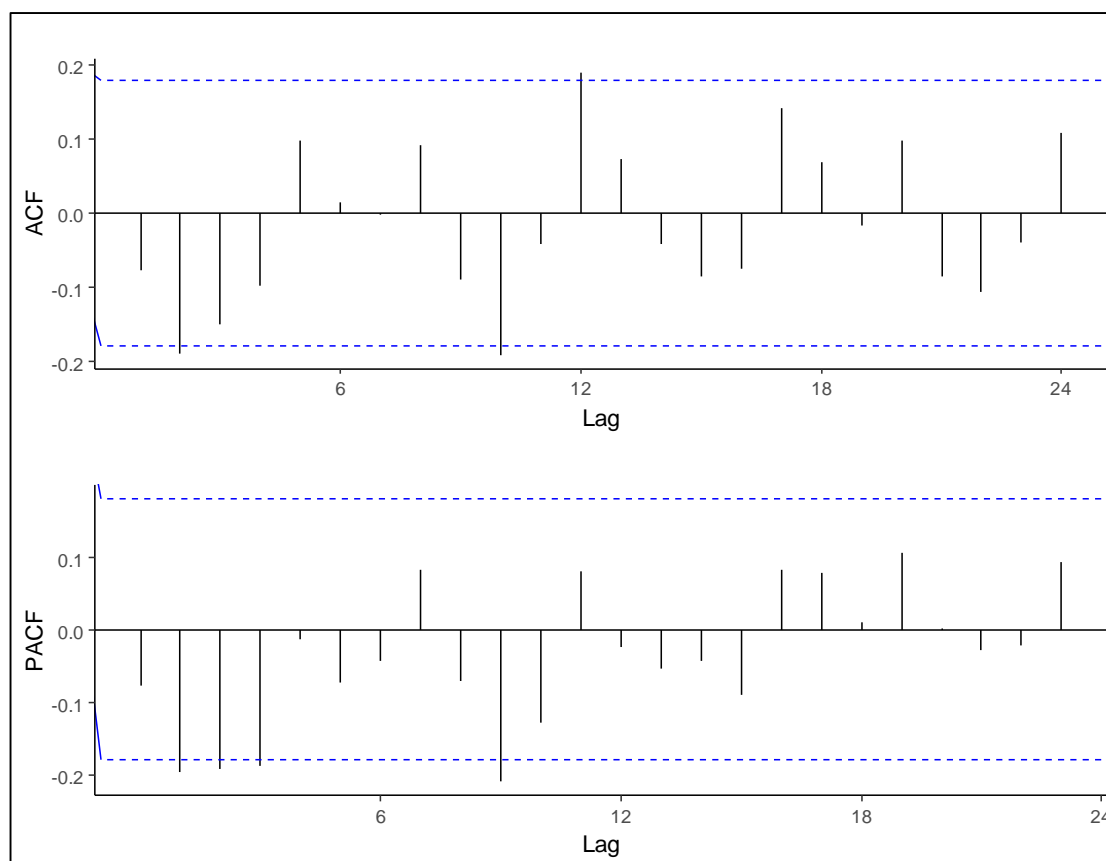


Figura 8: Correlograma FAC y FACP de la primera diferencia regular de la serie del número de pacientes atendidos, periodo 2008-2017.

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

Prueba analítica de Estacionariedad Dickey-Fuller para la primera diferencia regular de la serie del número de pacientes atendidos.

$H_0: p > 0.05$ La serie es no estacionaria (Tiene raíz unitaria)

$H_a: p < 0.05$ La serie es estacionaria (No tiene raíz unitaria)

Se aplicó el test de Dickey-Fuller y se obtuvo el valor de -6.3785 con un p -valor $= 0.01$ que es menor a nuestro nivel de significancia de 0.05 , por ende se rechaza la hipótesis nula y se acepta la hipótesis alterna.

Se concluye que la primera diferencia regular de la serie del número de pacientes atendidos es estacionaria.

Siguiendo la metodología Box-Jenkins, se presenta los resultados de la serie de la estimación, comprobación del diagnóstico y la predicción, con el Software estadístico R.

La elección del mejor modelo de predicción se muestra en la siguiente página. El modelo que presenta el menor índice de Akaike.

Ecuación

De los discutido anteriormente el modelo identificado es un ARIMA (1.1.1)(0.1.1)₁₂, cuya ecuación es la siguiente:

$$\hat{Y}_t = -0.07 * Y_{t-1} - 0.99 * Y_{t-2} - 0.78 * Y_{t-3} + Y_{t-12} - Y_{t-13} + \varepsilon_t - \varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}.$$

4.1.2. Fase de estimación del modelo identificado: arima (1.1.1)(0.1.1)₁₂

Continuando con la metodología de Box-Jenkins el paso a seguir es la especificación del modelo, en el mismo que se analizara la significancia tanto individual como global de sus parámetros.

Tabla 2: Comparativo de modelos estimados de la serie del número pacientes atendidos, periodo 2008-2017.

Modelo: ARIMA(1.0.1)(0.0.1) ₁₂				Modelo: ARIMA(1.1.1)(0.1.1) ₁₂			
	Coefficiente	z	Valor p		Coefficiente	z	Valor p
AR(1)	-0.679068	9.711	2e-16***	AR(1)	-0.079360	-0.800	0.4237
MA(1)	-0.99999	-41.0607	2e-16***	MA(1)	-0.999999	-27.3052	2.2e-16 ***
SMA(1)	0.215970	2.3697	0.0178*	SAR(1)	-0.783921	-6.3139	2.629e-10***
Criterio de Akaike	1775.17			Criterio de Akaike	1609.28		

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

Con la utilización del Software R utilizando la función Autoarima el cual sería un buen modelo según el software y como investigar propongo otro modelo, como se observa en la Tabla N°2 los criterios de información Akaike disminuyen de 1775.17 a 1609.28; esto significa que la diferencia entre la serie y la serie estimada es mínima.

Donde logramos encontrar que el modelo propuesto por el investigador tiene menor índice de Akaike y lo elegimos como el que mejor se ajusta a la serie.

El que mejor describe a la serie analizada es el modelo ARIMA (1.1.1)(0.1.1)_[12], con ecuación:

$$\hat{Y}_t = -0.07 * Y_{t-1} - 0.99 * Y_{t-2} - 0.78 * Y_{t-3} + Y_{t-12} - Y_{t-13} + \varepsilon_t - \varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}.$$

4.1.3. Validación del modelo

Después de seleccionar un modelo ARIMA particular y de estimar sus parámetros, se trata de ver si el modelo seleccionado se ajusta a los datos en forma razonablemente buena, el detalle se encuentra en ver la medida en que los residuos del modelo estimado se aproximan al comportamiento de un ruido blanco.

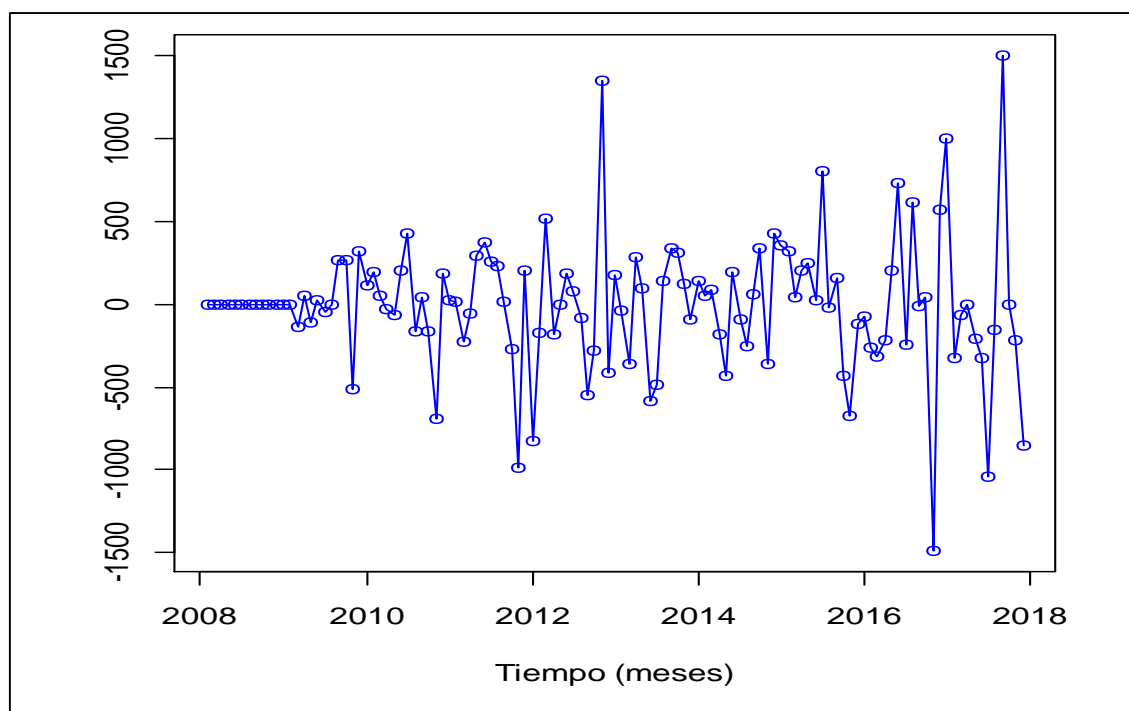


Figura 9: Residuales del modelo estimado de la serie del número de pacientes.

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

Como podemos ver en la Figura 9 los residuales tienen el comportamiento de un ruido blanco.

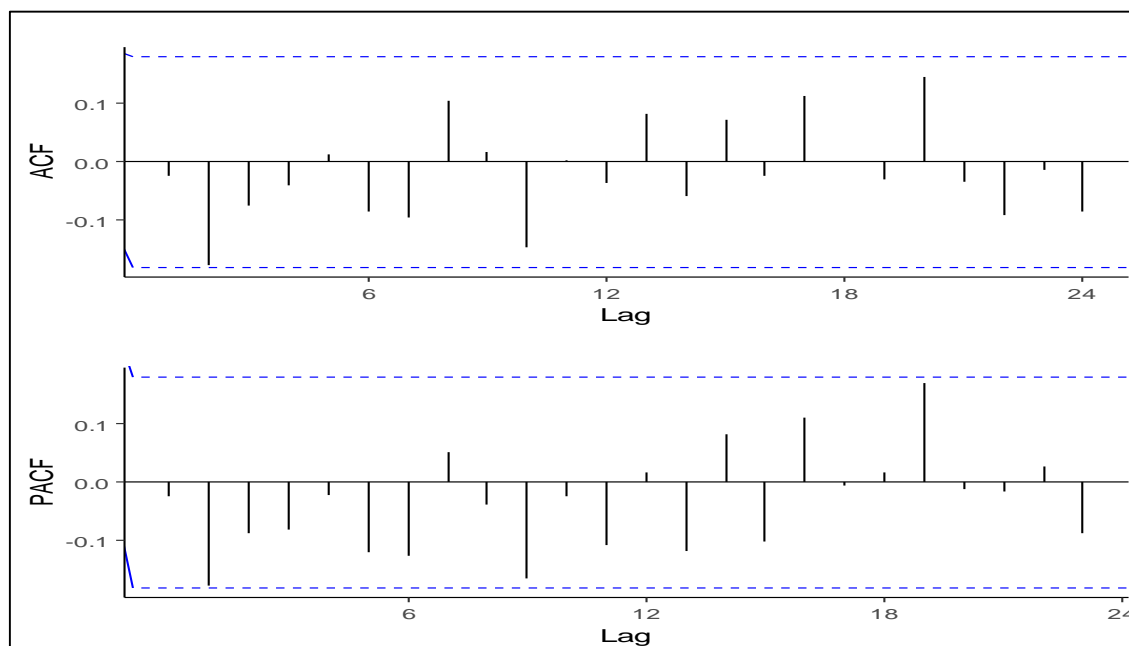


Figura 10: Correlograma FAC y FACP de los residuales del modelo de la serie del número de pacientes atendidos, periodo 2008-2017.

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

Para la propuesta de un nivel de significancia $\alpha = 0.05$ por lo menos el 95% de las correlaciones de los residuales toma valores dentro de las bandas, por lo que podemos admitir que los residuos que hemos obtenido se comportan como un ruido blanco.

Análisis de los residuos (ruido blanco)

La interpretación del estadístico Q Ljung - box es más favorable al ruido blanco cuanto mayor sea la probabilidad “p”, si se utiliza al nivel de confianza habitual del 95%, entonces los ruidos son ruido blanco siempre que el p-valor sea superior a 0.05, para el coeficiente de autocorrelación de que los residuos forman una secuencia aleatoria se realizó las siguientes comparaciones:

H0: $p > 0.05$ Los residuales son aleatorios (ruido blanco).

Ha: $p < 0.05$ Los residuales no son aleatorios.

Tabla 3: Prueba Ljung-Box Q' de la serie del número de pacientes atendidos.

Retardos	Ljung-Box Q'	Valor p
12	11.373	0.1814
24	21.073	0.3928
36	25.198	0.617

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

Estadístico Q de Ljung-Box con retrasos de 12, 24 y 36, los residuales son ruido blanco puesto que los valores de p para los retardos son superiores a 0.05, aceptamos la hipótesis nula.

Prueba de la Normalidad de los Residuales

H0: $p > 0.05$ Los residuales se distribuyen según una distribución normal

Ha: $p < 0.05$ Los residuales no se distribuyen según una distribución normal

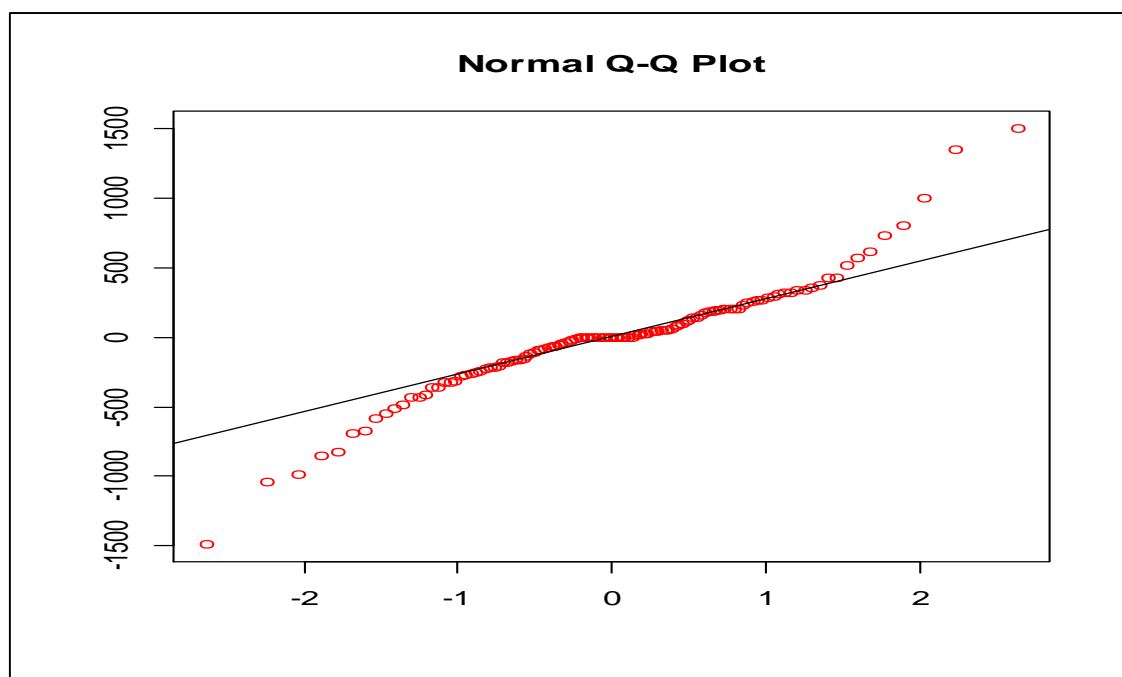


Figura 11: Grafico Q para evaluar la normalidad de los residuales del modelo.

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

El diagrama Q-Q de los residuos del modelo estimado para la serie se muestra en la Figura 11. Los puntos parecen seguir la línea recta bastante de cerca, aunque parece desviarse en el extremo inferior. Este gráfico no nos llevaría a rechazar la normalidad de los términos de error en este modelo.

Además, la prueba de normalidad de Shapiro-Wilk aplicada a los residuos produce un estadístico de prueba de $W = 0.92421$, lo que corresponde a un valor de p de 0.1379, y no rechazaría la normalidad basado en esta prueba.

Concluyendo que los residuos del modelo siguen una distribución normal es decir que los valores de los residuales de la serie giran en torno a su valor medio y confirmamos que el modelo ARIMA (1.1.1)(0.1.1)₁₂, con ecuación de pronóstico:

$$\hat{Y}_t = -0.07 * Y_{t-1} - 0.99 * Y_{t-2} - 0.78 * Y_{t-3} + Y_{t-12} - Y_{t-13} + \varepsilon_t - \varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}.$$

4.1.4. Realización del pronóstico con el modelo

La fase de predicción nos permite obtener resultados a futuro. La información proyectada es de 12 meses, desde la información existente del número de pacientes atendidos en el Hospital Regional Manuel Núñez Butrón - Puno.

Tabla 4: Pronóstico de la serie del número de pacientes.

Pronósticos con el Modelo: ARIMA(1.1.1)(0.1.1) _[12]				
$\hat{Y}_t = -0.07 * Y_{t-1} - 0.99 * Y_{t-2} - 0.78 * Y_{t-3} + Y_{t-12} - Y_{t-13} + \varepsilon_t - \varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}$				
Para intervalos de confianza 95%, $z(0.025) = 1.96$				
Año	Mes	Número de pacientes	Intervalo de confianza	
			Inferior	Superior
2018	Enero	630	609.537	651.417
	Febrero	1439	1417.9	1459.78
	Marzo	1743	1722.155	1764.035
	Abril	1687	1666.914	1708.794
	Mayo	1450	1429.696	1471.576
	Junio	1350	1329.395	1371.275
	Julio	1433	1412.485	1454.365
	Agosto	1811	1790.286	1832.166
	Septiembre	2067	2046.363	2088.243
	Octubre	1909	1887.709	1929.589
	Noviembre	1543	1521.721	1563.601
	Diciembre	1414	1392.717	1434.597

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

En esta fase de predicción nos permite obtener resultados a futuro. La información proyectada es de 12 meses, desde la información existente del número de pacientes atendidos en el Hospital Regional Manuel Núñez Butrón – Puno.

Los valores previstos para el número de pacientes atendidos en el Hospital Regional Manuel Núñez Butrón – Puno, durante el periodo donde los datos reales están disponibles, también se muestran los valores predichos a partir del modelo ajustado y los residuos. Para los periodos de tiempo más allá del final de la serie, se muestra 95% límites de predicción para las previsiones. Estos límites muestran donde es probable encontrar los verdaderos, valores en un momento futuro seleccionado al 95% de confianza, asumiendo que el modelo ajustado es apropiado para los datos.

Para las observaciones

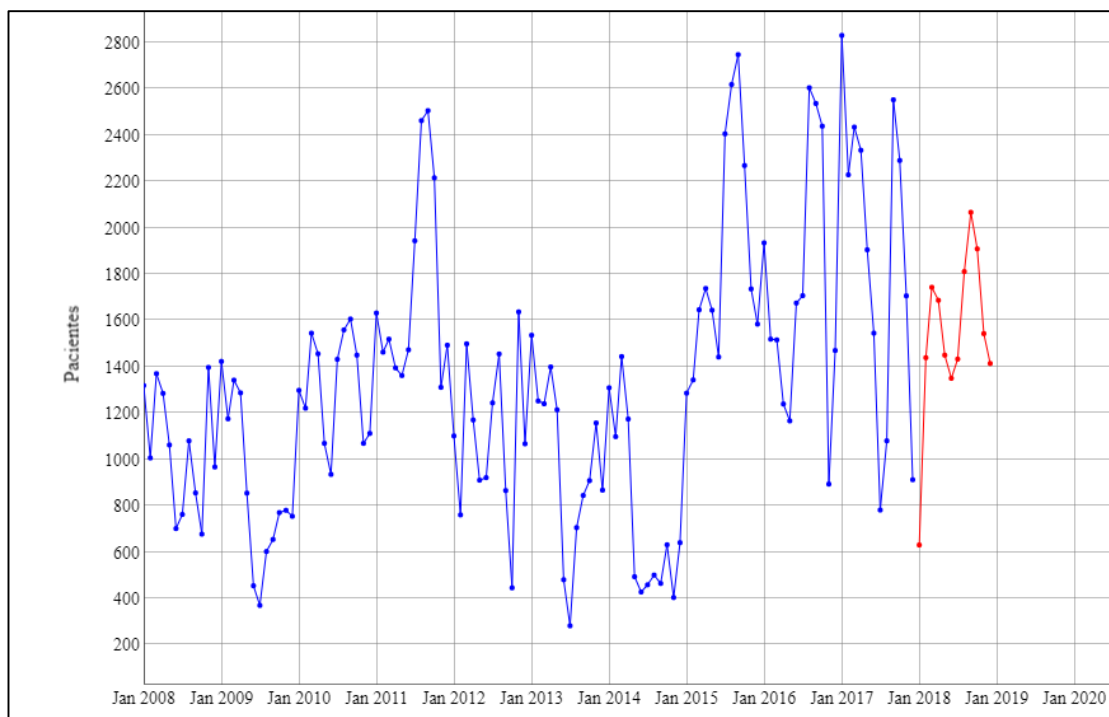


Figura 12: Predicción del número de pacientes atendidos para el año 2018.

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

En la Figura 12 se muestran la predicción del número de pacientes atendidos en el Hospital Regional Manuel Núñez Butrón – Puno proyectada para 12 meses del modelo ARIMA (1.1.1)(0.1.1)_[12], con ecuación: $\hat{Y}_t = -0.07 * Y_{t-1} - 0.99 * Y_{t-2} - 0.78 * Y_{t-3} + Y_{t-12} - Y_{t-13} + \varepsilon_t - \varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}$

CONCLUSIONES

PRIMERA:

El modelo que se determinó utilizando la tendencia de la serie número de pacientes del Hospital Regional Manuel Núñez Butrón es el modelo ARIMA.

SEGUNDA:

Los parámetros del modelo que mejor se ajustan al número de pacientes en el Hospital Regional Manuel Núñez Butrón para el año 2018 son ARIMA (1.1.1)(0.1.1)_[12], con su ecuación de pronóstico estimada es:

$$\hat{Y}_t = -0.07 * Y_{t-1} - 0.99 * Y_{t-2} - 0.78 * Y_{t-3} + Y_{t-12} - Y_{t-13} + \varepsilon_t - \varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}.$$

TERCERA:

Al validar el modelo ARIMA(1.1.1)(0.1.1) que mejor se ajusta para el número de pacientes del Hospital Regional Manuel Núñez Butrón vemos claramente que los residuales tienen un comportamiento de un ruido blanco, sumado a esto el índice de Akaike de nuestro modelo propuesto es menor al del modelo propuesto por el software R.

CUARTA:

Se obtuvieron las proyecciones mensuales para el año 2018, de la serie de número de pacientes del Hospital Regional Manuel Núñez Butrón, para enero=630, febrero=1439, marzo=1743, abril=1687, mayo=1450, junio=1350, julio=1433, agosto 1811, septiembre=2067, octubre=1909, noviembre=1543, diciembre=1414, con su respectiva amplitud de intervalo de predicción al 95% de confianza.

RECOMENDACIONES

PRIMERA:

Para realizar una buena estimación de un modelo AR, MA y/o ARIMA, se requiere de una serie temporal de tamaño grande para la variable mínimo 54 observaciones

SEGUNDA:

Tomar como referencia trabajos similares al efectuado para mayor entendimiento del tema.

TERCERA:

Se recomienda realizar trabajos de investigación del presente estudio cada cierto periodo de tiempo.

CUARTA:

A nuevos investigadores, experimentar con casos nuevos para corroborar lo expuesto anteriormente y experimentar con nuevas metodologías y esto permita hacer comparaciones entre diferentes metodologías y así mejorar las predicciones en trabajos futuros.

REFERENCIAS BIBLIOGRÁFICAS

- Anderson, D. (1985). *"Time Series Analysis and Forecasting"*. (1ra Edición) Londres: Editions Lonchers Polytechnics y Butterworths
- Aznar, A. y Trivez, F. (1993). *"Métodos de predicción en Economía II. Análisis de series de temporales"*. Barcelona: Ariel Economía.
- Aznar, A. García, A. y Martín, A. (1994). *"Ejercicios de Econometría I"*. Madrid: Ediciones Pirámide S.A.
- Bedoya, S. (2018). *"Modelamiento univariado del número de defunciones infantiles producidas por infecciones respiratorias agudas, a través de la Metodología Box-Jenkins, Puno 2008-2016"*
- Espaza, A. y Cancelo, J. (1993). *"Métodos cuantitativos para el análisis de la coyuntura Económica"*. Madrid: Alianza.
- Guerrero, M. (2003). *"Análisis Estadístico de Series de Tiempo Económicas"*. México: Thomson Editores.
- Hernández, R., Fernández, C. y Baptista, P. (2010). *"Metodología de la investigación"*. México: McGRAW-HILL.
- Ibañez, V. (1998). *"Análisis de Series Temporales y Pronóstico"*. Puno-Perú.
- Melo, E. (2016). *"Modelo de predicción mensual de mortalidad general intrahospitalaria en el Hospital Regional Manuel Núñez Butrón, Puno 2008-2016 I"*
- Merlín, W. (2015). *"Modelo univariante de pronóstico del número de unidades de transfusión de sangre en el Hospital regional Manuel Núñez Butrón, Puno. periodo 2006-2015 I"*

- O'Connell, R., Bowerman, L. y koehler, B. (2007). *"Pronósticos, Series de Tiempo y Regresión"*. México: Thomson Editores. Paraninfo.
- Pindyck, S., y Rubinfeld, L. (2003). *"Econometría, Modelos y Pronósticos"*. Madrid: McGraw-Hill.
- Quispe, J. (2008). *"Modelo de predicción de los nacimientos y defunciones del distrito de juliaca-2008"*.
- Quispe, Y. (2015). *"Modelo univariante para el consumo de energía eléctrica doméstica en el Distrito de Ayaviri – Electro Puno, periodo 2004-2013"*
- Uriel, E. (1985). *"Análisis de Series Temporales - Modelos Arima"*. Madrid: Thomson Editores
- Valdez, Y. (2015). *"Modelo de predicción mensual del número de intervenciones quirúrgicas más frecuentes en el Hospital Regional Manuel Núñez Butrón – Puno 2006 – 2015-I"*

WEBGRAFÍA

- http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D
- <http://docplayer.es/181816-Modelos-de-series-de-tiempo-para-predecir-la-inflacion-en-venezuela.html>
- http://portal.uc3m.es/portal/page/portal/dpto_estadistica/personal/isabel_molina_peralta
- <https://addi.ehu.es/bitstream/10810/12492/1/04-09gon.pdf>
- https://es.wikipedia.org/wiki/Metodolog%C3%ADa_de_Box-Jenkins

ANEXOS

Anexo 1: Datos originales del número de atenciones según meses del periodo 2008 – 2017

	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	
MEDICINA INTERNA	ENERO	59	78	51	61	30	88	68	33	46	210
	FEBRERO	74	74	97	105	65	82	66	127	18	318
	MARZO	64	93	90	51	75	53	132	48	177	386
	ABRIL	66	67	121	110	76	55	78	164	243	291
	MAYO	55	18	64	78	77	32	4	124	118	385
	JUNIO	29	19	71	61	48	9	29	113	180	257
	JULIO	54	33	101	81	95	13	54	149	238	17
	AGOSTO	55	35	132	113	93	16	55	210	314	167
	SETIEMBRE	49	39	124	114	70	28	49	199	302	469
	OCTUBRE	26	35	61	48	30	22	48	73	298	408
	NOVIEMBRE	42	46	85	88	35	49	42	127	57	275
	DICIEMBRE	33	41	86	97	14	52	29	142	121	75
GINECOLOGIA Y OBSTETRICIA	ENERO	456	497	466	414	350	561	432	500	327	704
	FEBRERO	335	429	442	439	171	498	360	524	353	582
	MARZO	530	442	593	521	727	332	552	633	408	608
	ABRIL	445	472	557	493	414	476	468	646	339	533
	MAYO	333	335	460	452	226	439	231	688	215	138
	JUNIO	202	139	298	425	214	190	88	508	341	21
	JULIO	303	74	444	499	529	77	71	817	181	240
	AGOSTO	448	204	459	564	559	336	71	846	282	66
	SETIEMBRE	301	247	432	572	275	326	168	695	448	453
	OCTUBRE	297	344	431	547	171	423	265	596	497	438
	NOVIEMBRE	609	259	267	339	779	439	79	455	222	425
	DICIEMBRE	415	278	332	423	532	297	258	406	440	484
CIRUGIA	ENERO	137	155	178	173	142	131	179	176	170	153
	FEBRERO	86	126	157	125	70	101	151	163	86	175
	MARZO	136	137	153	138	154	118	156	149	127	183
	ABRIL	112	110	159	166	105	119	101	216	116	172
	MAYO	109	93	91	140	73	145	41	140	139	155
	JUNIO	68	49	110	162	104	31	68	153	171	153
	JULIO	84	62	127	150	127	40	84	170	130	18
	AGOSTO	93	71	132	174	137	49	93	170	177	85
	SETIEMBRE	93	95	111	157	90	96	93	128	185	187
	OCTUBRE	57	60	78	174	31	83	36	119	228	177
	NOVIEMBRE	137	123	118	70	163	110	137	99	41	186
	DICIEMBRE	114	72	49	73	99	128	15	83	63	27
MEDICINA FISICA Y	ENERO	176	157	43	482	72	279	35	51	912	1345
	FEBRERO	78	82	25	360	31	125	39	11	709	758
	MARZO	96	102	116	343	43	149	54	177	508	766
	ABRIL	154	114	152	293	134	173	54	250	335	918
	MAYO	91	78	120	268	50	132	23	216	319	791

	JUNIO	34	36	83	337	31	37	34	131	542	705
	JULIO	31	34	395	701	24	37	31	759	643	471
	AGOSTO	77	57	470	1054	117	37	77	863	1244	594
	SETIEMBRE	38	37	575	1089	40	36	38	1112	1066	973
	OCTUBRE	35	37	550	1045	15	54	20	1080	1009	810
	NOVIEMBRE	122	51	363	557	169	74	28	698	416	459
	DICIEMBRE	44	13	313	617	79	9	16	610	624	158
PEDIATRIA	ENERO	230	279	304	277	223	236	321	287	267	191
	FEBRERO	206	236	265	224	183	229	242	288	159	170
	MARZO	262	285	273	214	229	294	276	269	159	207
	ABRIL	249	255	220	157	218	280	230	209	105	225
	MAYO	294	193	187	246	307	281	104	269	223	238
	JUNIO	216	87	170	267	313	118	56	284	249	219
	JULIO	126	36	162	280	232	19	53	271	288	15
	AGOSTO	236	102	136	260	300	172	32	239	280	83
	SETIEMBRE	197	120	152	243	212	182	58	245	240	196
	OCTUBRE	145	150	156	212	109	181	118	193	231	216
	NOVIEMBRE	255	142	105	119	262	248	35	174	63	186
	DICIEMBRE	193	175	152	116	189	196	153	150	81	73
OFTALMOLOGIA	ENERO	263	258	257	226	284	241	274	239	213	228
	FEBRERO	229	229	235	212	240	217	240	230	194	226
	MARZO	283	284	322	254	271	294	274	370	137	285
	ABRIL	260	270	248	177	223	296	243	253	101	196
	MAYO	181	138	149	180	177	185	90	207	152	198
	JUNIO	153	124	203	223	211	95	153	253	192	190
	JULIO	166	131	203	234	237	95	166	240	227	20
	AGOSTO	172	134	232	300	249	95	172	291	308	85
	SETIEMBRE	177	117	214	333	178	176	58	369	296	275
	OCTUBRE	117	145	176	192	89	145	144	208	176	242
	NOVIEMBRE	233	160	133	139	229	237	82	183	94	175
	DICIEMBRE	170	177	181	167	154	185	169	193	141	95

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

Anexo 2: primera diferencia de los datos del número de pacientes

	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dec
2008		-313	364	-85	-223	-361	61	318	-225	-178	720	-430
2009	456	-248	167	-55	-433	-400	-85	233	52	116	10	-26
2010	544	-77	324	-89	-387	-134	497	127	47	-156	-381	43
2011	520	-169	56	-124	-33	111	471	519	43	-290	-905	182
2012	-392	-341	739	-329	-260	11	323	211	-590	-420	1192	-570
2013	469	-248	-12	159	-185	-734	-199	424	139	64	249	-290
2014	442	-211	346	-270	-681	-66	31	42	-36	167	-228	237
2015	646	57	303	92	-94	-202	964	213	129	-479	-533	-152
2016	351	-416	-3	-277	-73	509	32	898	-68	-98	-1546	577
2017	1361	-602	206	-100	-430	-360	-764	299	1473	-262	-585	-794

FUENTE: Elaboración propia en base a datos del HRMNB-PUNO.

Anexo 3: Residuo de los datos del número de pacientes

	Ene	Feb	Mar	Abr	May
2008		-1,807105E+05	4,435853E+05	-8,907914E+04	-1,881002E+05
2009	6,017719E+05	5,999260E+03	-1,398976E+08	5,429333E+07	-1,126806E+08
2010	1,142094E+08	1,948035E+08	5,319096E+07	-3,171144E+07	-6,363102E+07
2011	2,042859E+07	1,728269E+07	-2,222917E+08	-5,667866E+07	2,931000E+08
2012	-8,305657E+08	-1,758485E+08	5,158526E+08	-1,813773E+08	-6,307630E+06
2013	1,786390E+08	-3,666586E+07	-3,576717E+08	2,862744E+08	9,443238E+07
2014	1,382234E+08	5,159715E+07	8,664440E+07	-1,788811E+08	-4,346431E+08
2015	3,522828E+08	3,179714E+08	3,756143E+07	2,056560E+08	2,482678E+08
2016	-7,479226E+07	-2,591937E+08	-3,123163E+08	-2,214668E+08	2,007603E+08
2017	1,001122E+09	-3,230743E+08	-6,648610E+07	-1,096707E+06	-2,054537E+08

	Jun	Jul	Ago	Sep	Oct
2008	-2,748776E+05	1,594712E+05	3,758168E+05	-1,814170E+05	-1,167358E+05
2009	2,016077E+07	-5,059623E+07	-4,387088E+06	2,659097E+08	2,678493E+08
2010	1,986260E+08	4,237339E+08	-1,638395E+08	4,394531E+07	-1,643765E+08
2011	3,727881E+08	2,577108E+08	2,326292E+08	1,604690E+07	-2,714326E+08
2012	1,865234E+08	7,667538E+07	-8,765861E+07	-5,491740E+08	-2,791012E+08
2013	-5,843088E+08	-4,826499E+08	1,408958E+08	3,392305E+08	3,131860E+08
2014	1,908991E+08	-9,291917E+07	-2,574761E+08	6,077434E+07	3,354681E+08
2015	2,211082E+07	7,980398E+08	-1,707629E+07	1,576495E+08	-4,285470E+08
2016	7,348986E+08	-2,420787E+08	6,105857E+08	-1,010424E+07	3,809345E+07
2017	-3,276848E+08	-1,037420E+09	-1,553640E+08	1,503621E+09	-4,823628E+06

	Nov	Dic
2008	7,491364E+05	-4,179770E+05
2009	-5,088762E+08	3,162408E+08
2010	-6,961579E+08	1,807740E+08
2011	-9,912836E+08	2,018230E+08
2012	1,350977E+09	-4,178481E+08
2013	1,190473E+08	-9,621139E+07
2014	-3,579293E+08	4,252105E+08
2015	-6,738308E+08	-1,163364E+08
2016	-1,489971E+09	5,697039E+08
2017	-2,181129E+08	-8,489262E+08