

UNIVERSIDAD NACIONAL DEL ALTIPLANO - PUNO

FACULTAD DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA

ESCUELA PROFESIONAL DE INGENIERÍA ESTADÍSTICA E INFORMÁTICA



**MODELO UNIVARIANTE DE PRONÓSTICO DEL NÚMERO DE
UNIDADES DE TRANSFUSIÓN DE SANGRE EN EL HOSPITAL
REGIONAL MANUEL NUÑEZ BUTRÓN - PUNO PERIODO 2006-**

2015-I

TESIS

PRESENTADA POR:

Bach. WALTER OCTAVIO MERLIN MAMANI

PARA OPTAR EL TÍTULO PROFESIONAL DE:

INGENIERO ESTADÍSTICO E INFORMÁTICO

PUNO – PERÚ

2015



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



TESIS

MODELO UNIVARIANTE DE PRONÓSTICO DEL NÚMERO DE
UNIDADES DE TRANSFUSIÓN DE SANGRE EN EL HOSPITAL
REGIONAL MANUEL NUÑEZ BUTRÓN - PUNO PERIODO 2006-
2015-I

PRESENTADA POR:

Bach. WALTER OCTAVIO MERLIN MAMANI

A la Coordinación de Investigación de la Facultad de Ingeniería Estadística e
Informática de la Universidad Nacional del Altiplano - Puno, para optar el Título
Profesional de:

INGENIERO ESTADÍSTICO E INFORMÁTICO

APROBADA POR:


PRESIDENTE

: 
Dr. VLADIMIRO IBAÑEZ QUISPE

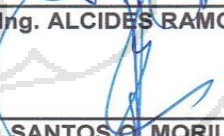
PRIMER MIEMBRO

: 
M.Sc. ALEJANDRO APAZA TARQUI

SEGUNDO MIEMBRO

: 
Ing. ALCIDES RAMOS CALCINA

DIRECTOR

: 
M.C. SANTOS O. MORILLOS VALDERRAMA

ASESOR

: 
Mg. EMMA O. AZAÑERO DE AGUIRRE

ÁREA: Estadística
TEMA: Series de tiempo

DEDICATORIAS

A Jaent Estrella por guiar y alentarme cada pasó de mi vida por el camino del bien y otorgarme el don de vivir y hacer posible la realización de mi meta.

Con mucho cariño e infinita gratitud a mis queridos padres Basilio y Margarita, quienes me ofrecieron su constante apoyo y aliento para seguir adelante en todo aquello que emprendía.

A mis hermanos(as) por el respaldo moral que me brindaron a cada instante con inmenso cariño incondicional y por la esperanza que me dan al saber que superarán mis logros.

Walter Merlin

AGRADECIMIENTOS

- ☞ A la Universidad Nacional del Altiplano – Puno, y al Facultad de Ingeniería de Estadística e Informática; en especial a todos los docentes de la Escuela Profesional de Ingeniería de Estadística e Informática por nuestra formación Profesional.
- ☞ Al Hospital Regional Manuel Núñez Butrón de Puno, en especial a mi Asesor y Director que me brindaron muchos conocimientos en nuestro formación Profesional.
- ☞ A nuestros compañeros y amigos presentes y pasados, quienes sin esperar nada a cambio compartieron sus conocimientos, alegrías, tristezas con nosotros, y a todas aquellas personas que durante estos años de estudio estuvieron apoyándonos y lograron que este sueño se haga realidad.

ÍNDICE

DEDICATORIAS	i
AGRADECIMIENTOS	ii
ÍNDICE DE TABLAS	v
ÍNDICE DE FIGURAS	vi
ÍNDICE DE GRÁFICOS	vii
RESUMEN..	viii
INTRODUCCIÓN	x

CAPÍTULO I

PLAN DE INVESTIGACIÓN

1.1. PROBLEMA	1
1.2. JUSTIFICACIÓN DE LA INVESTIGACIÓN	3
1.3. OBJETIVOS DE LA INVESTIGACIÓN	4
1.3.1. OBJETIVO GENERAL.....	4
1.3.2. OBJETIVOS ESPECÍFICOS	4
1.4. HIPÓTESIS	4
1.5. LIMITACIONES.....	5

CAPÍTULO II

MARCO TEORICO

2.1. ANTECEDENTES DE LA INVESTIGACIÓN	6
2.2. BASE TEÓRICA.....	7
2.3. DEFINICIÓN DE TÉRMINOS BÁSICOS	44
2.4. OPERACIONALIZACIÓN DE VARIABLES	49

CAPÍTULO III**MATERIALES Y MÉTODOS**

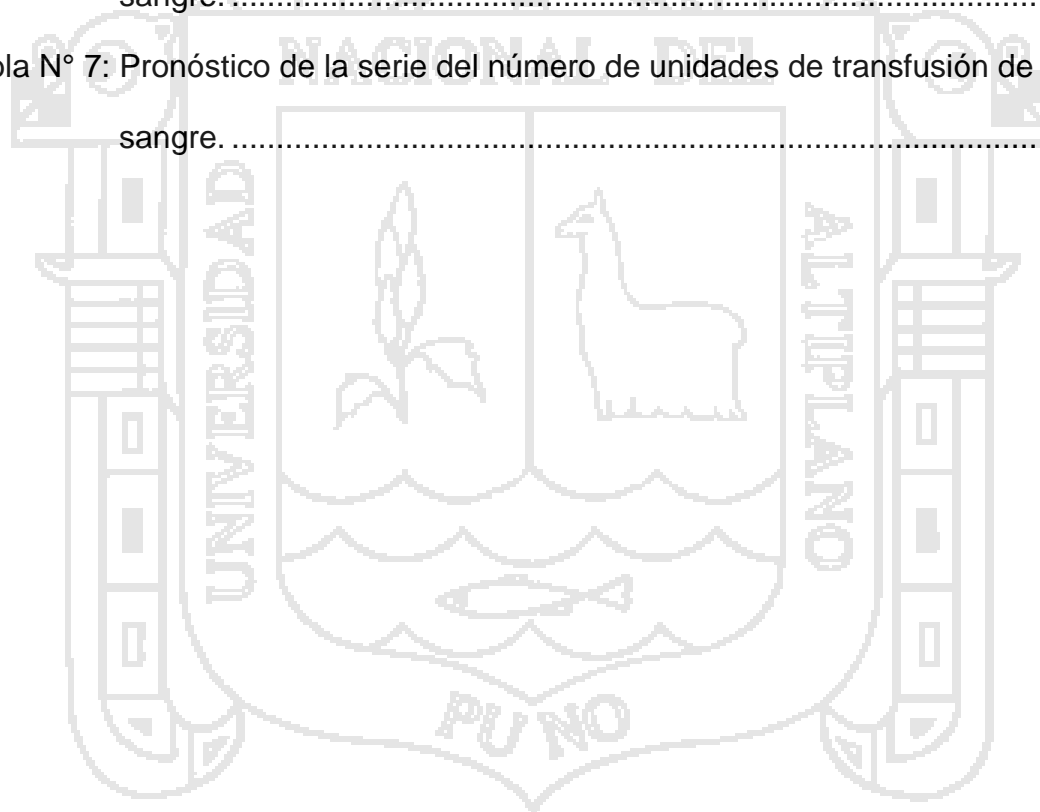
3.1. LOCALIZACIÓN.....	50
3.2. MATERIALES Y MÉTODOS	50
3.3. POBLACIÓN	50
3.4. MUESTRA.....	51
3.5. UNIDAD MUESTRAL	51
3.6. MÉTODOS DE RECOPIACIÓN DE DATOS.....	51
3.7. METODOLOGÍA	51

CAPÍTULO IV**RESULTADOS Y DISCUSIÓN**

4.1. ANALISIS ESTADÍSTICO DE LA SERIE TIEMPO	59
4.1.1. FASE DE IDENTIFICACIÓN DEL MODELO	59
4.1.2. FASE DE ESTIMACIÓN DEL MODELO	67
4.1.3. FASE DE EVALUACIÓN O VALIDACIÓN DEL MODELO	69
4.1.4. FASE DE REALIZACIÓN DE PRONÓSTICO CON EL MODELO... 73	
CONCLUSIONES	76
RECOMENDACIONES Y SUGERENCIAS	77
BIBLIOGRAFÍA	78
REFERENCIAS DE TESIS	79
CITAS ELECTRÓNICAS	80
ANEXOS.....	81

ÍNDICE DE TABLAS

Tabla N° 1: Comportamiento de las FAC y FACP.....	35
Tabla N° 2: Tipos de Sangre y Compatibilidad.	43
Tabla N° 3: Operacionalizacion de Variables.....	49
Tabla N° 4: Prueba de raíz unitaria para la serie del número de unidades de transfusión de sangre.....	69
Tabla N° 5: Comparativo de modelos estimados de la serie del número de unidades de transfusión de sangre.	68
Tabla N° 6: Prueba de Ljung-Box Q' de la serie del número de transfusiones de sangre.	70
Tabla N° 7: Pronóstico de la serie del número de unidades de transfusión de sangre.	75



ÍNDICE DE FIGURAS

Figura N° 1: Función de Autocorrelación y Autocorrelación Parcial.....	35
Figura N° 2 : Metodología del Enfoque Box- Jenkins.....	53



ÍNDICE DE GRÁFICOS

Gráfico N° 1: número de unidades de transfusiones de sangre, periodo 2006 - 2015-I.....	59
Gráfico N° 2: Histograma de Serie de tiempo del número de Unidades de transfusiones de sangre.....	60
Gráfico N° 3: Serie del número de transfusiones de sangre por años	61
Gráfico N° 4: Media y Varianza del número de unidades de transfusiones de sangre por años.	62
Gráfico N° 5: Correlogramas de FAC y FACP de la serie del número de transfusiones de sangre.....	63
Gráfico N° 6: Primera diferencia regular de la serie del número de unidades de transfusión de sangre.....	65
Gráfico N° 7: Primera diferencia regular y estacional de la serie del número de unidades de transfusión de sangre.	66
Gráfico N° 8: Normalidad de la serie del número de unidades de transfusión de sangre transformadas.	67
Gráfico N° 9: Q-Qplot Cuartiles normales de la serie del número de unidades de transfusión de sangre transformadas.....	68
Gráfico N°10: Correlogramas de FAC y FACP de la serie del número de unidades de transfusión de sangre transformadas.	70
Gráfico N°11: Prueba de normalidad de los residuos para la serie del número de unidades de transfusión de sangre.	71
Gráfico N°12: FAC Y FACP de los residuos para la serie del número de transfusiones de sangre.....	72
Gráfico N°13: Serie del número de unidades de transfusión de sangre y la serie de pronóstico.....	74

RESUMEN

El Hospital Regional Manuel Núñez Butrón Puno, está ubicado a 3819 m.s.n.m de la provincia de Puno, orientado exclusivamente a la actividad de servicio de recuperación y rehabilitación, en los pacientes por ende cuenta con profesionales en diferente áreas del servicio, como el banco de sangre, necesitando futuras proyecciones del comportamiento del número de unidades de transfusión de sangre ya que son datos de suma importancia para programar un determinado presupuesto para el año siguiente, para así prever recursos humanos, infraestructura, equipamiento tecnológicos y financieros de tal manera que la presente investigación tiene como objetivo determinar el modelo de predicción mensual de ajuste del número de unidades de transfusión de sangre en el Hospital Regional “Manuel Núñez Butrón” Puno, periodo 2006-2015-I.

Se hizo uso de la Metodología Box- Jenkins para el modelamiento de la serie del número de unidades de transfusión de sangre, la cual se desarrolló en las siguientes etapas exploración de la serie, identificación de modelo, estimación de los parámetros del modelo tentativo, verificación de modelo y finalmente usar el modelo tentativo para el pronóstico.

El resultado de trabajo de investigación el mejor modelo univariante para pronosticar la serie del número de unidades de transfusión de sangre, en el Hospital Regional Manuel Núñez Butrón - Puno es un modelo **SARIMA** (0, 1,1) (1, 1,0)¹²

Palabras Claves: Modelo, estimar, parámetros, univariante.

ABSTRACT

The Regional Hospital Manuel Nunez Butron Puno, is located 3819 meters above sea level in the province of Puno, aimed exclusively at the service activity recovery and rehabilitation, patients therefore has professionals in different areas of the service, as the blood bank , necessitating future projections of the behavior of the number of units of blood transfusion because they are important data adds to schedule a budget for next year so provide human resources, infrastructure, technology and financial equipment so that this investigation It aims to determine the model of predicting monthly adjustment of the number of units of blood transfusion in the Regional Hospital "Manuel Núñez Butrón" Puno period 2006-2015-I.

Was made, using the Box-Jenkins methodology for modeling of the series the number of units of blood transfusion, which took place on the steps of exploration of the series, model identification, parameter estimation of the tentative model, model verification and finally use the tentative model for forecasting.

The result of research the best univariate model to predict the number the number of units of blood transfusion, in the Regional Hospital Manuel Nunez Butron - Puno is a SARIMA (0, 1,1) model (1, 1,0)¹².

Keywords: Model, estimating parameters, univariate.

INTRODUCCIÓN

Uno de los problemas que carece el Ministerio de Salud, es que aún no existen trabajos de investigación que describan su comportamiento en el futuro. Por cuanto a estudios de los modelos de series de tiempo constituye una necesidad para describir y efectuar las predicciones que estos ayuden para prever recursos humanos, infraestructura, equipamiento tecnológico y financiero para la toma de decisiones a futuro. Por lo tanto el estudio de series de tiempo, dentro de ellos los modelos univariantes constituye una necesidad para efectuar predicciones a futuro.

La importancia de saber el comportamiento de las variaciones del número de unidades de transfusiones de sangre derivado en un futuro es muy importante para la prevención y planificación ante ciertos incidentes que pueden surgir producto de pacientes que tienen la anemia o cuando los niveles de hemoglobina son demasiado bajos, hoy en día las transfusiones sanguíneas innecesarias aumentan el costo de la atención médica y escasez relativa de donaciones de sangre.

Y concuerdo con la opinión de (Centeno, 2013) *“la importancia para el hospital de Apoyo de sandía es conocer pronósticos y así poder prever un determinado presupuesto para el año siguiente”*.

El presente trabajo de investigación identificar el modelo tentativo y posteriormente realizar el pronóstico para el número de unidades de transfusión de sangre, empleando la metodología de Box Jenkins. En los siguientes capítulos se muestra todo los procesos que ameritan para el modelo de

pronóstico del número de unidades de transfusión de sangre del Hospital Regional “Manuel Núñez Butrón” Puno, cada capítulo se detalla a continuación:

En el capítulo I, se explica los fundamentos para la realización de la tesis, así como planteamiento 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, definición de términos básicos, operacionalización de variables.

En el capítulo III, se describe los materiales y métodos para el modelo de pronóstico, para el número de unidades de transfusión de sangre en el Hospital Regional Manuel Núñez Butrón Puno.

En el capítulo IV, se muestra los resultados que el mejor modelo de ajuste para el número de transfusiones de sangre en el Hospital Regional Manuel Núñez Butrón Puno periodo 2006 – 2015-I, es **SARIMA** (0, 1,1) (1, 1,0)₁₂, cuya ecuación de pronóstico es: $\hat{Y}_t = Y_{t-1} + Y_{t-2} - 0.35Y_{t-12} - 0.35Y_{t-13} + \varepsilon_t + 0.89\varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}$

Por último se muestra, las conclusiones y recomendaciones sobre el modelo univariante integrado de Box-Jenkins proporciona un mejor ajuste para validar y predecir el número de unidades de transfusión de sangre en el en el Hospital Regional Manuel Núñez Butrón - Puno.

CAPÍTULO I

PLAN DE INVESTIGACIÓN

1.1. PROBLEMA

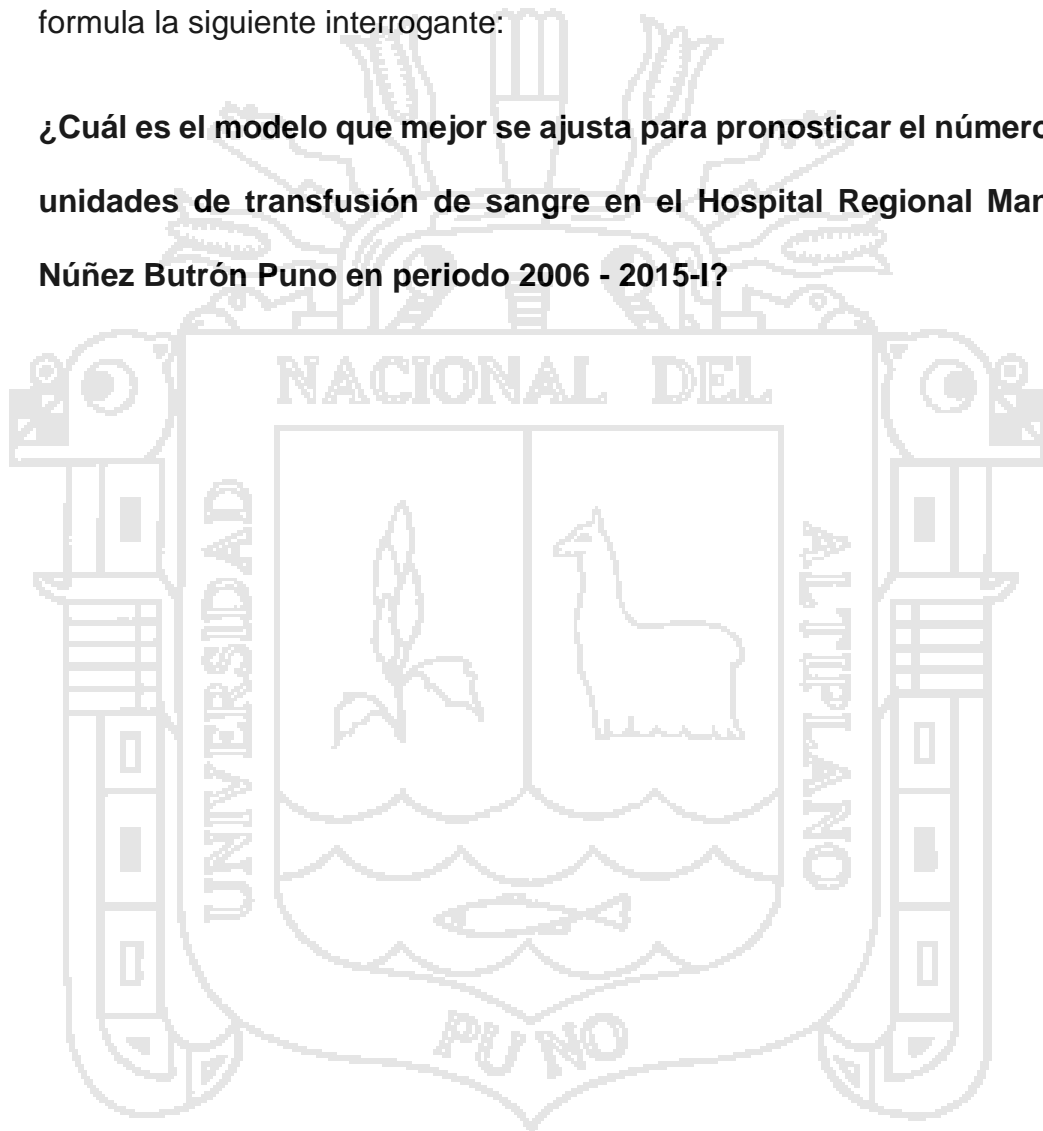
El Hospital Regional Manuel Núñez Butrón de Puno cuenta con Programa Nacional de Hemoterapia y Bancos de Sangre que presenta una serie de deficiencias como:

La Unidad de Estadística e Informática, no cuenta con suficiente información sobre el número de unidades transfusión de sangre. No existe un registro de control, ya que a veces el Director del Hospital Regional Manuel Núñez Butrón pide información para tomar decisiones adecuadas, sin embargo resulta que esto no se cumple, por parte del Órgano Institucional de Salud.

Otro alcance del problema es que, en los últimos años, la población de la provincia de Puno, tiene un crecimiento continuo, esto implica la mejora de la oferta y calidad del servicio de salud en beneficio de la población, siendo necesario realizar predicciones a futuro, esto conlleva a que el HRMNB puno prevé los servicios de banco de sangre que llegan a ocurrir en el transcurso de los años, para luego obtener resultados muy confiables.

Frente a esta problemática de vital importancia, para la población de la Región Puno, con el deseo de contribuir con el conocimiento para el análisis y la toma de decisiones hacia el futuro, con el fin de buscar una solución inteligente al problema y dar alternativas de prevención a los problemas prioritarios del Programa Nacional de Hemoterapia y Bancos de Sangre, se formula la siguiente interrogante:

¿Cuál es el modelo que mejor se ajusta para pronosticar el número de unidades de transfusión de sangre en el Hospital Regional Manuel Núñez Butrón Puno en periodo 2006 - 2015-I?



1.2. JUSTIFICACIÓN DE LA INVESTIGACIÓN

La presente investigación se realiza por la creciente dependencia de calidad de servicio de banco sangre. Por lo tanto la presente investigación beneficiara al Hospital Regional Manuel Núñez Butrón Puno, en donde permitirá conocer a futuro el comportamiento del número de unidades de transfusión de sangre y así prever las necesidades con el servicio que brinda el Hospital Regional Manuel Núñez Butrón Puno a la población de la Región de Puno.

Una de las técnicas estadísticas más empleadas en series de tiempo es usando la metodología de Box-Jenkins, de manera que permita tomar acertadas decisiones, y permite obtener buenas aproximaciones cuando se hace un buen uso de las series de tiempo estacionarias o que puedan convertirse en estacionarias mediante la transformación adecuada.

Por parte del Estado tenemos, la Ley 26454 que creó el Programa Nacional de Hemoterapia y Bancos de Sangre (PRONAHEBAS) como órgano competente del Ministerio de Salud, dicha ley en su Capítulo V, de la Donación y Transfusión de Sangre, el Art. 18 dice: La Transfusión de sangre humana, sus componentes y derivados con fines terapéuticos, constituye un acto médico que debe llevarse a cabo únicamente después de un estudio racional y específico de la patología a tratar, lo cual indica establecer los mecanismos adecuados para su uso debidamente justificado (MINSA,2014).

1.3. OBJETIVOS DE LA INVESTIGACIÓN

1.3.1.OBJETIVO GENERAL

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.

1.3.2.OBJETIVOS ESPECÍFICOS

Identificar, estimar y validar con el modelo identificado que se ajusta para predecir el número de unidades de transfusión de sangre en el Hospital Regional Manuel Núñez Butrón Puno en el periodo 2006- 2015-I.

Realizar el pronóstico con el modelo alcanzado, para el número de unidades de transfusión de sangre en el Hospital Regional Manuel Núñez Butrón Puno en el periodo 2006-2015-I.

1.4. HIPÓTESIS

El modelo SARIMA proporciona un mejor ajuste para realizar el pronóstico del número de unidades transfusión de sangre en el Hospital Regional Manuel Núñez Butrón Puno 2006 – 2015-I.

1.5. LIMITACIONES

Una de las limitaciones que se tuvo para la presente investigación es la recopilación de información, es decir que no se tiene datos completos anteriores a los años 2006 en la Unidad de Estadística e Informática y en banco de sangre en el Hospital Regional Manuel Núñez Butrón Puno.

Otra limitación es el licenciamiento del software que se usó en la presente investigación.



CAPÍTULO II

MARCO TEÓRICO

2.1. ANTECEDENTES DE LA INVESTIGACIÓN

Carvajal, M. (2011), concluye con el mejor modelo que se ajusta para realizar predicciones de transfusiones sanguíneas en los pacientes ingresados al Hospital Verdi Cevallos en España es ARIMA (1, 0,0).

Cutipa, P. (2010), Reporta que el mejor modelo que se ajusta para predecirlas Atenciones por Emergencia en la Provincia de Puno, según servicio es: ARIMA(6,1,6) para la serie de atenciones de Medicina General, ARIMA(1,2,1) para la serie de atenciones de Cirugía General, ARIMA(1,2,1) para la serie de atenciones de Pediatría, ARIMA(1,2,1) para la serie de atenciones de Gineco-Obstetiz.

Centeno, D. (2013), reporta que el mejor modelo que se ajusta para estimar el proceso de atenciones en el Hospital de apoyo Sandia, es SARIMA (1, 0,0)*(0, 1,1).

Valdez, T. (2015), concluye con el mejor modelo que se ajusta para pronosticar las Intervenciones quirúrgicas; “aborto” en la Hospital Regional Manuel Núñez- Puno, es SARIMA (1, 1,1) (0, 1,1).

2.2. BASE TEÓRICA

BREVE RESEÑA HISTÓRICA

La historia de transfusión sanguínea empezó, cuando la primera investigación en la transfusión de sangre data del siglo XVII en que el médico Británico Guillermo Harvey describió completo la circulación y las propiedades de la sangre en 1628, las transfusiones de primera sangre también tentativa alrededor de este tiempo, aunque éstos fueran a menudo fatales fracasado y probado en seres humanos.

En 1667, Jean Baptiste Denis que era médico a Rey Louis XIV, realizado la transfusión de la sangre de un animal a un ser humano. Denis hizo una trasfusión la sangre de una oveja a un muchacho de 15 años y más adelante a un trabajador, ambos de los cuales sobrevivieron las transfusiones.

Entre 1914 y 1918, los anticoagulantes tales como citrato de sodio fueron encontrados para prolongar la vida útil de la sangre y de la refrigeración también demostradas ser los medios efectivos de preservar sangre. *En los años 20 y los años 30, la donación voluntaria de la sangre para el almacenamiento y el uso fueron comenzados, durante la Segunda Guerra Mundial, la transfusión de sangre fue utilizada a gran escala para tratar a soldados heridos y llegó a ser bien conocida como procedimiento de la salvación de vidas (Sally R, 2014).*

SERIE DE TIEMPO

Una serie tiempo es una secuencia de observaciones, medidos en determinados momentos del tiempo, ordenados cronológicamente y, espaciados entre sí de manera uniforme, así los datos usualmente son dependientes entre sí (Olivera, 1985).

La serie de tiempo es una información básica de la evolución de variables en el tiempo. Entre ellos los modelos de Box-Jenkins constituyen un conjunto de procedimiento para el tratamiento de serie de tiempo. (Grass & Ferrer, 1994).

Es un conjunto de observaciones ordenadas según una característica cuantitativa de un fenómeno individual, en diferentes momentos de tiempo (Hanke, 1996).

TÉCNICAS DE PREDICCIÓN

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 (Hanke, 1996).

COMPONENTES DE UNA SERIE TEMPORAL

LA TENDENCIA

Es un componente de una serie temporal que refleja su evolución a largo plazo. Puede ser de naturaleza estacionaria o constante (se representa con una recta paralela al eje de las abscisas), de naturaleza lineal, parabólica, exponencial (Hamilton, 1994).

LAS VARIACIONES CÍCLICAS

Es un componente de la serie que recoge oscilaciones periódicas de amplitud superior a un año. Estas oscilaciones periódicas no son regulares y se presentan en los fenómenos económicos cuando se dan de forma alternativa, etapas de prosperidad o de depresión (Anderson, 1995).

LAS VARIACIONES ESTACIONALES

Es un componente de la serie que recoge oscilaciones que se producen alrededor de la tendencia, de forma repetitiva y en periodos iguales o inferiores a un año. Por ejemplo, el clima afecta a la venta de una serie de productos, los helados y refrescos se venden fundamentalmente en verano

y la ropa de abrigo en invierno, entre los factores más importantes que originan variaciones estacionales, se encuentran las condiciones climáticas, las costumbres sociales y las fiestas religiosas (Hamilton, 1994).

LAS VARIACIONES ACCIDENTALES Y/O IRREGULARES

(Aznar & Trivez, 1993) menciona que: *“Es una componente de la serie que recoge movimientos provocados por factores imprevisibles (un pedido inesperado a nuestra empresa, una huelga, también se conoce con el nombre de variaciones irregulares, residuales o erráticas)”*.

CLASIFICACIÓN DESCRIPTIVA DE LAS SERIES TEMPORALES

Estacionarias

Una serie es estacionaria cuando es estable a lo largo del tiempo, es decir, cuando la media y varianza son constantes en el tiempo. Esto se refleja gráficamente en que los valores de la serie tienden a oscilar alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo (Uriel, 1985).

No estacionarias

Son series en las cuales la tendencia y/o variabilidad cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante (Uriel, 1985).

MODELO

Un modelo es una expresión formalizada de una teoría, o la representación matemática de los datos observados. En el análisis estadístico un modelo es expresado, en símbolos de forma matemática. Para la construcción de un buen modelo es necesario contar con el conjunto de datos observados. *También es importante la experiencia, la intuición, la imaginación, la simplicidad y la habilidad para seleccionar el subconjunto más pequeño de variables. El primer paso es establecer el problema en forma clara y lógica delimitando sus fronteras, luego viene la recogida y la depuración de datos, el diseño del experimento; las pruebas de contraste; la verificación del modelo y la validación de las hipótesis (Andrew, 1994).*

Un modelo debe ser una buena aproximación al sistema real, debe incorporar los aspectos importantes del sistema y debe resultar fácil de comprender y manejar. Un factor muy importante es que debe presentar una alta correlación entre lo que predice el modelo y lo que actualmente ocurre en el sistema real (Andrew, 1994).

MODELO DE SERIES TEMPORALES

Son formas teóricas determinísticas y/o aleatorias o la combinación de ambas, para realizar el análisis de una serie de tiempo.

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

Serie Temporal: Es el conjunto de “ t ” observaciones, una observación por cada una de las variables: Y_1, Y_2, \dots, Y_t . también es denominada serie cronológica.

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 singularidad 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 constituye un mecanismo explicativo que recibe el nombre del modelo (Fernández, 1970).

Procesos estocásticos estacionarios

Se denomina proceso estocástico a la sucesión infinita de variables aleatorias ordenadas.

$$y_1, y_2, y_3, \dots, y_t$$

Si se dispone de un conjunto finito de esas variables, $y_1, y_2, y_3, \dots, y_t$, 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 ε_t se denomina "ruido blanco" si cumple las siguientes condiciones:

- i. $E(\varepsilon_t) = 0 \quad \forall t$
- ii. $V(\varepsilon_t) = E(\varepsilon_t^2) = \sigma^2 \quad \forall t$
- iii. $COV(\varepsilon_t) = E(\varepsilon_t, \varepsilon_s) = 0 \quad \forall t \neq s$

Esto es, la variable Y_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 X_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 siguiente expresión:

$$a_t \sim Niid(0, \sigma^2) \quad \forall t$$

Niid:: Normales independientes e idénticamente distribuidas

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 (Uriel, 1985).

Proceso estocástico y Estacionariedad

Los modelos autorregresivos o de medias móviles que más adelante conceptualizaremos necesitan para su comprensión de la introducción del concepto de proceso estocástico.

Un proceso estocástico es una sucesión de variables aleatorias y_t ordenadas, pudiendo tomar t cualquier valor entre $-\infty$ y $+\infty$, por ejemplo, la siguiente sucesión de variables aleatorias puede ser considerado como proceso estocástico.

$$\dots, Y_1, Y_2, \dots, Y_T, \dots$$

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

Cada una de las variables y_t , que configuran un proceso estocástico tendrá su propia función de distribución de 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 ocurría, ya no para cada par de variables, si no 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.

$$Y_{t1}, Y_{t2}, Y_{t3}, \dots, Y_{tm}$$

Cualesquiera que fueran los valores de $(Y_{t_1}, Y_{t_2}, Y_{t_3}, \dots, Y_{t_m})$ 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 varianza para cada Y_t y la covarianza para variables referidas a distintos valores de t .

Procesos estacionarios en sentido amplio

Se dice que una serie temporal y_t ha sido generada por un proceso estocástico estacionario en sentido amplio si se cumplen simultáneamente las tres condiciones siguientes:

Cada observación tiende a oscilar alrededor de una media que es constante a lo largo del tiempo. Es decir, todas las variables del proceso tienen la misma esperanza matemática:

$$E(Y_t) = \mu \quad \forall t$$

La dispersión alrededor de esa media constante a lo largo del tiempo también es constante. Es decir, todas las variables del proceso tienen la misma varianza:

$$VAR(Y_t) = \sigma_y^2 \quad \forall t$$

La covarianza entre dos variables que disten k períodos de tiempo (autocovarianza de orden k) es la misma que existe entre cualesquiera otras dos variables que disten también k períodos entre sí, independientemente del momento del tiempo al que estén referidas:

$$\text{COV}(Y_t, Y_{t+k}) = \text{COV}(Y_{t+j}, Y_{t+j+k}) = \gamma_k \quad \forall t$$

Debe tenerse mucho cuidado en no confundir un proceso estacionario con un proceso que presenta estacionalidad (fluctuaciones regulares dentro del año) (Aznar & Trivez, 1993).

RUIDO BLANCO

El 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).

MODELO UNIVARIANTE

Es una serie de tiempo $\{Y_t\}$, los modelos univariantes se consideran todos aquellos que solamente tiene una sola variable observada en el tiempo. Estos tipos de modelos se expresan en forma polinomial.

Entre las técnicas univariantes existen algunas muy sencillas, tales como el modelo autorregresivo de primer orden, el modelo de tendencia lineal o exponencial, entre otros.

Las técnicas más rigurosas para la predicción univariante son las denominadas técnicas o modelos de Box-Jenkins, o más concretamente

modelos ARIMA, pues las técnicas de Box-Jenkins constituyen un conjunto más amplio, dentro del cual los modelos ARIMA univariantes son solo una parte (Aznar & Trivez, 1993).

MODELO UNIVARIANTE DE BOX-JENKINS NO- INTEGRADOS

Los procesos autorregresivo AR (p), las Medias móviles MA (q) y procesos mixtos ARMA (p, q) son considerados como los modelos No Integrados debido a que no interviene el grado de diferenciación y la estacionalidad de la serie observada.

MODELO UNIVARIANTE DE BOX-JINKINS INTEGRADOS

A los procesos mixtos integrados ARIMA (p, d, q), los procesos estacional mixto integrado SARIMA (p, d, q) x (P, D, Q), proceso de medias móviles exponenciales EWMA, y los procesos de auto agregación, se considera como modelos integrados por que intervine la estacionalidad de la serie en estudio.

ESTACIONALIDAD

Si no fuera por la estacionalidad, el análisis de las series temporales se convertiría en una tarea muy simple. De hecho , la mayoría de los series temporales, e ciencias psicológicas y sociales ,quedarían bien representadas por los modelos de orden inferior, tales como los modelos ARIMA(1,0,0), ARIMA(0,0,1) ARIMA(0,1,0), ARIMA(1,1,0) Y ARIMA(0,1,1). Estos modelos cubren gran parte de los procesos temporales fácilmente identificables. Por desgracia, suele ocurrir que las series temporales

presentan ciclos de carácter mensual, trimestral, anual, etc. De modo que se hallan afectados por fuertes componentes estacionales.

La estacionalidad se define por una fluctuación cíclica o periódica de la serie temporal que se repite de forma regular. Desde un punto de vista analítico, la estacionalidad constituye una variación del proceso de debe ser extraída o controlada. Una forma de control de la variación estacional consiste en desestacionar la serie antes su análisis.

El mejor procedimiento para el manejo de la estacionalidad es construir un modelo causal de las fuerzas estacionales. Es decir, tener en cuenta.

NO ESTACIONARIEDAD ESTACIONAL

Un proceso observado o realización empírica puede presentar una inclinación o tendencia en pasos o incrementos, por ejemplo anuales. Por dicha razón, a fin de tener en cuenta la inclinación o tendencia estacional, la serie debe ser diferenciada estacionalmente. Así con datos mensuales se sustrae Y_t de Y_{t-12} y Y_{t-1} de Y_{t-12} , etc. Este proceso se define, mediante el operador de diferencia, por:

$$(1 - \Delta^{12})y_t = \theta_0$$

$$y_t = y_{t-12} + \theta_0$$

MODELO ARIMA ESTACIONAL

A partir de lo expuesto en los puntos anteriores, es posible expresar el modelo estacional ARIMA, en términos generales, por ARIMA (p, d, q) (P, D, Q) S, donde P, D, Q, son análogos a los parámetros p, d, q. el parámetro S simboliza la longitud del periodo o ciclo. De esta forma. Con los datos mensuales $S=12$. Al modelizar procesos estacionales, solemos incorporar tanto estructuras regulares como como estacionales de forma aditiva.

MODELOS LINEALES ESTACIONARIOS

Modelos Autorregresivos (AR)

Definimos un modelo como autorregresivo si la variable endógena de un periodo t es explicada por las observaciones de ella misma correspondientes a periodos anteriores añadiéndose, como en los modelos estructurales, un término de error. En este caso de procesos estacionarios con distribución normal, la teoría estadística de procesos estocásticos afirma que, bajo determinadas condiciones previas, toda Y_t puede expresarse como una combinación lineal de sus valores pasados.

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 de tiempo analizadas que intervienen en la ecuación.

La expresión genérica de un modelo autorregresivo AR (p) sería de la siguiente forma:

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

Donde: ε_t es una variable “ruido blanco”, siempre y cuando cumpla lo siguiente:

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

Los procesos AR (p), también se puede escribir de forma abreviada como:

$$\phi_p(L)Y_t = \delta + \varepsilon_t$$

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

$$\phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3 - \dots - \phi_p L^p$$

Y donde, a su vez el termino L, es el que se conoce como operador retardo tal que, implicado 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 aplicando sucesivamente p veces retarda, el valor en p periodos.

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

Normalmente, se suele trabajar con modelos autorregresivos de ordenes bajos: AR (1), AR (2), o bien con órdenes coincidentes con la prioridad de datos de la serie analizada, si es trimestral AR (4), si es mensual AR (12). Si resulta que $p = 12$ para datos mensuales, el modelo autorregresivo establece un modelo de índices estacionales que son los coeficientes estimados. Como se mencionó previamente, puede eliminarse el propio patrón estacional para investigar si hay otro modelo que abarca varios años, o si el modelo se extiende a un plazo más largo. Naturalmente, el modelo autorregresivo puede también revelar variaciones cíclicas menores de doce meses. Se debe tener en cuenta que es necesario imponer ciertas restricciones a los valores de los parámetros de este modelo para que funcione correctamente estacionario

MODELO AR (1)

Un modelo AR (1) viene definido por:

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

Entonces utilizando el operador de retardos para un modelo AR (1) sería de la siguiente forma:

$$(1 - \phi_1 L)Y_t = \varepsilon_t$$

Donde ε_t es ruido blanco

MODELO AR (2)

Un modelo AR (2) viene definido por:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$$

Así mismo aplicando para un modelo AR (2), su operador de retardos sería de la siguiente forma:

$$(1 - \phi_1 L - \phi_2 L^2) Y_t = \varepsilon_t$$

PROCESOS DE MEDIAS MÓVILES (MA).

Un modelo de Box-Jenkins denominados de medias móviles es aquel que explica el valor de una determinada variable en un periodo 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. La notación, MA (q) que se refiere a un modelo de media móvil de orden q . Entonces la expresión genérica de un modelo autorregresivo MA (q) sería de la siguiente forma:

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

Donde a_t es un ruido blanco con las propiedades ya definidas.

Obsérvese que el proceso de medias móviles corresponde a una combinación lineal de variables ruido blanco, siendo los coeficientes "theta"

los “ponderadores” de la combinación lineal. Esto es, se asemeja a la definición de un “promedio” de las variables ruido blanco (aunque los coeficientes no sumen en general la unidad, una propiedad que deben cumplir los promedios).

MODELO MA (1)

Un modelo MA (1) viene definido por:

$$Y_t = a_t - \theta_1 a_{t-1} = (1 - \theta_1 L) a_t$$

Donde: a_t es un ruido blanco con las propiedades ya definidas.

MODELO MA (2)

Un modelo MA (2) viene definido por:

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} = (1 - \theta_1 L - \theta_2 L^2) a_t$$

Donde: a_t es un ruido blanco.

Proceso Autorregresivo de Media Móvil (ARMA)

La combinación de procesos AR y MA da lugar a los procesos mixtos ARMA. La notación de ARMA (p, q), se refiere a un modelo con p términos autorregresivo y q términos de media móvil.

La formulación general de un proceso ARMA, ARMA (p, q), es:

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

Los momentos del proceso se derivan de la misma forma que se vio para los procesos AR y MA.

Condiciones y Raíces Unitaria para los Modelos

Hemos afirmado anteriormente que, bajo condiciones generales de todo el proceso estocástico se presenta una especificación tipo AR (p) y en consecuencia podía expresarse también como un MA (q). Es ahora el momento para 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, aquí estudiaremos que deben de cumplirse dos condiciones importantes.

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 esta misma tome en $t+k$, siendo k 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 en el tiempo del momento para el que estamos considerando dicha correlación.

La Estacionariedad de las Series Temporales en la Realidad

En términos generales vemos que a, aquellos procesos estocásticos que cumplan, al menos de forma débil, la restricción de la 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 importantes.

Primeramente demos asegurarnos de que la serie temporal, como muestra del proceso estocástico, es estacionaria y, si no lo es. Hay que transformar la serie temporal original de tal 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.**

Para resolver la primera cuestión existen diversos métodos de aproximación y, de entre ellos, destacamos Podríamos subdividir la serie temporal en varios períodos de, aproximadamente, 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 $\hat{r}_t = y_t - \hat{T}_t$, 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

$$\text{ARIMA } \hat{y}_t = \hat{T}_t + \hat{r}_t.$$

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). 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:

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

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

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

Se trata por tanto de contrastar si el coeficiente a_1 es igual a la unidad o distinto de ella.

Sin embargo, para contrastar la nulidad del coeficiente a_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 ($a_1=0$) de la (Ec.2), sin embargo, en nuestro caso, necesitaríamos contrastar $H_0: a_1 = 1$. 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:

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

La estimación de a_1 en $y_t = a_1 y_{t-1} + \varepsilon_t$ 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, por cuestiones de sencillez operativa, el modelo utilizado para el contraste DF no es el expuesto al comienzo del epígrafe sino otro, equivalente al anterior, que se obtiene restando a uno y otro lado el término

y_{t-1} :

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

$$\Delta y_t = a_0 + (a_1 - 1)y_{t-1} + \varepsilon_t \quad (\text{Ec. 2})$$

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

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 $a_1 = 1$ 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 $a_1 > 1$.

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.

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. Por eso propusieron, entre otros autores, seguir un proceso en etapas a fin de garantizar el éxito en la elección de un modelo (O'Connell, Bowerman & Koehler, 2007).

PROCESOS ARIMA- NO ESTACIONARIOS

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” (modelos ARMA), es necesario discutir mecanismos de transformación de las series a procesos estacionarios.

En 1970, Box y Jenkins desarrollaron un cuerpo metodológico destinado a identificar, estimar y diagnosticar modelos dinámicos de series temporales en los que la variable tiempo juega un papel fundamental.

En principio pueden presentarse distintas (infinitas) formas por las que se introduce la no estacionariedad en un proceso estocástico. Sin embargo, interesa considerar solo algunas formas de la no estacionariedad que sean adecuados para describir el comportamiento de series temporales y, al mismo tiempo, posibles de ser transformados en procesos estacionarios.

MODELO ARIMA (p, d, q) SARIMA (P, D, Q)

Generalmente se suele expresar como ARIMA (p, d, q) donde los parámetros p, d y q son números enteros no negativos que indican el orden de las distintas componentes del modelo respectivamente, las

componentes autorregresivo, integrada y de media móvil. Cuando alguno de los tres parámetros es cero, es común omitir las letras correspondientes del acrónimo AR para la componente autorregresivo, I para la integrada y MA para la media móvil.

En su forma más general el modelo ARIMA (p, d, q) ARIMA (P, D, Q,)s podría escribirse como:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_{P_S+p+D_S+d} Y_{t-P_S-p-sD-d} + \delta + U_t + \theta_1 U_{t-1} + \dots + \theta_{Q_S+q} U_{t-sQ-q}$$

Entendiendo que puede haber más de un proceso generador de la serie (en la parte regular y en la estacional) y escribiendo una combinación de los modelos MA (q) y AR (p) que han precisado de una serie de diferenciaciones "d" en la parte regular o "D" en la parte estacional para que fueran 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 estuvo perfectamente definido si conociéramos su media constante (μ), su varianza constante (σ) y la covarianza entre cada par de momentos diferentes en el tiempo.

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:

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

Donde, evidentemente, cuando el valor de "k" 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:

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

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

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad \forall k$$

Por las propiedades de la covarianza y la definición del coeficiente de correlación, se tiene:

$$\rho_k = \rho_{-k} \quad \forall k$$

$$\rho_0 = \frac{\gamma_0}{\gamma_0} \equiv 1$$

El conjunto (infinito) de los valores de los coeficientes de autocorrelación se denomina función de autocorrelación (FAC), que es una función en j, cuyos valores son las autocorrelaciones definidas.

FUNCIÓN DE AUTO CORRELACIÓN

La función de autocorrelación (fac) y la función de autocorrelación parcial (facp) miden la relación estadística entre las observaciones de una serie temporal. Por ejemplo, el coeficiente de autocorrelación entre la variable y_t y la misma variable un período antes, y_{t-1} al que denominaremos coeficiente de autocorrelación de primer orden, se formula como:

$$\rho_1 = \frac{Cov(y_t, y_{t-1})}{\sqrt{var(y_t)}\sqrt{var(y_{t-1})}}$$

Dado el supuesto de estacionariedad, se tiene que $Var(y_t) = Var(y_{t-1})$, por lo que

$$\rho_1 = \frac{Cov(y_t, y_{t-1})}{Var(y_t)}$$

En general, para un desfase de k períodos se tiene que:

$$\rho_k = \frac{Cov(y_t, y_{t-k})}{Var(y_t)}$$

Y cuando $k=0$,

$$\rho_0 = \frac{Cov(y_t, y_t)}{Var(y_t)} = \frac{Var(y_t)}{Var(y_t)} = 1$$

FUNCIÓN DE AUTOCORRELACIÓN PARCIAL

La función de autocorrelación parcial mide la “aportación” que a las variaciones de una variable como y_t tiene otra variable, digamos y_{t-2} aislados los efectos de las posibles restantes variables, por ejemplo y_{t-1} . Por el contrario, la función de autocorrelación ignora el hecho de que parte de la correlación que pueda existir entre, por ejemplo y_t y y_{t-2} , se debe a que ambas están correlacionadas con y_{t-1} . Pues bien, los distintos coeficientes de autocorrelación parcial de los modelos teóricos se denotan como ϕ_{kk} , y los estimados para nuestra muestra como $\hat{\phi}_{kk}$.

Esta función entre el rango $[-1, +1]$, donde $+1$ indica una correlación perfecta (la señal se superpone perfectamente tras un desplazamiento temporal de K) y -1 indica una anticorrelación perfecta.

Función de autocorrelación parcial, con el fin de tener en cuenta los valores de correlación entre dos variables aleatorias separadas entre sí " k " períodos y en función de los valores intermedios entre ellas.

Es decir:

$$\Pi_k = \text{corr}(y_t, y_{t-k} : y_{t-1}, y_{t-2}, \dots, y_{t-k+1})$$

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

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

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

$$\Pi_k = \frac{\text{Cov}[(y_t - \hat{y}_t)(y_{t-k} - \hat{y}_{t-k})]}{\sqrt{\text{var}(y_t - \hat{y}_t)}\sqrt{\text{var}(y_{t-k} - \hat{y}_{t-k})}}$$

Pudiendo demostrarse que:

$$\Pi_k = \frac{\rho_k - \alpha_1 \rho_{k-1} - \alpha_2 \rho_{k-2} - \dots - \alpha_{k-1} \rho_1}{1 - \alpha_1 \rho_1 - \alpha_2 \rho_2 - \dots - \alpha_{k-1} \rho_{k-1}}$$

Figura N° 1: Función de Autocorrelación y Autocorrelación Parcial.

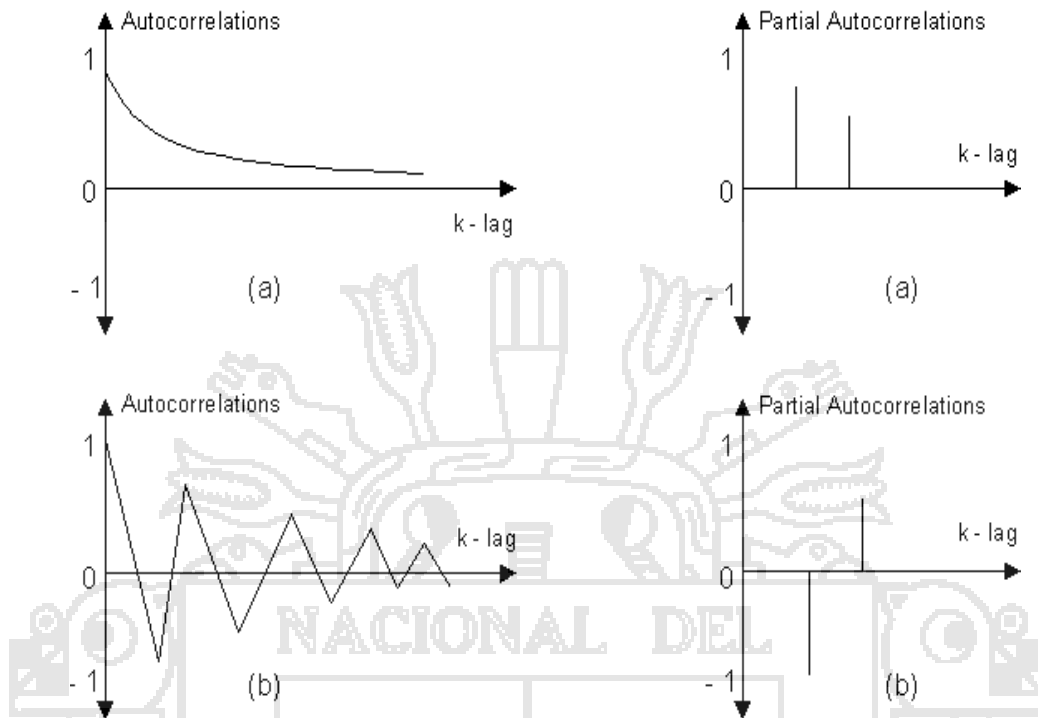


Tabla N° 1: Comportamiento de las FAC y FACP.

	FAC	FAP
MA(q)	se anulan para retardos superiores a q	Decrecimiento rápido sin llegar a anularse
AR(p)	Decrecimiento rápido sin llegar a anularse	se anulan para retardos superiores a p
ARMA(p, q)	Decrecimiento rápido sin llegar a anularse	Decrecimiento rápido sin llegar a anularse

HETEROCEDASTICIDAD

Una fuente importante de no estacionariedad corresponde a la presencia de heterocedasticidad en una serie dada. En ciertos casos ello puede corregirse con la aplicación de logaritmos (si la serie presenta tendencia creciente, es muy probable que la varianza de los valores originales también sea creciente en el tiempo).

PROCESOS ARIMA ESTACIONALES

Series de tiempo de frecuencia menor al año (mensuales, trimestrales) pueden presentar estacionalidad. Es decir, son series con ciclos u oscilaciones estrictamente periódicas, donde el período es igual o inferior al año. La presencia de este componente se explica por la existencia de las estaciones y su impacto sobre la actividad económica (por ejemplo, en la producción agropecuaria, el turismo), las costumbres (el fin de año, que es estrictamente cultural), procesos físicos (temperatura, pluviosidad), etc.

TRANSFORMACIÓN DE BOX-COX

Box-Cox (1964) definieron una transformación instantánea en el sentido de que no está involucrado simultáneamente varios periodos de tiempo de carácter más general de la transformación logarítmica. Esta transformación se define por:

$$Y_t^\lambda = \begin{cases} (Y_t^\lambda - 1)/\lambda, & \lambda \neq 0 \\ \ln y_t, & \lambda = 0 \end{cases}$$

La transformación de Box-Cox requiere definir el parámetro λ de la transformación.

Cuando el parámetro es $\lambda = 1$, la transformación de Box-Cox consiste prácticamente en tomar logaritmos.

Para el siguiente trabajo de investigación de utilizo la transformación por diferencias, para que la serie sea estacionaria.

La primera igualdad vale también, en el límite, de las diferencias de la serie original.

DESCRIPCIÓN TEÓRICA DE LOS PROCEDIMIENTOS DE DETECCIÓN DE RAÍCES UNITARIAS

(Universidad Autónoma de Madrid, 2015) *la aplicación de la metodología ARIMA precisa la utilización de series de tiempo estacionarias en varianza y en media. Cuando una serie no es estacionaria en media, o lo que es lo mismo, cuando no es integrada de orden cero I (0), se dice que presenta al menos una raíz unitaria. Cuando esto ocurre, sabemos que es posible la obtención de una serie estacionaria mediante una sencilla transformación de la serie original, como es la diferenciación adecuada. Pues bien, el número de diferencias que habrá que tomar en la serie para convertirla en estacionaria en media viene dado, justamente, por el número de raíces unitarias que la serie original. Pues bien, el número de raíces unitarias de la serie se determina formalmente mediante los procedimientos de Dickey-Fuller (con su versión ampliada) y de Phillips Perron.*

El planteamiento más sencillo de Dickey y Fuller es el siguiente. Plantean un modelo autorregresivo AR (1) como:

$$y_t = \mu + \rho y_{t-1} + \varepsilon_t$$

Donde μ y ρ son parámetros a estimar y ε_t es un término de error para el que se asume que cumple las propiedades de ruido blanco. Dado que ρ es un coeficiente de autocorrelación que, por tanto, toma valores entre $-1 < \rho < 1$, si $\rho = 1$, la serie y es no estacionaria. Si el valor de ρ es mayor que la unidad, entonces se dice que la serie es explosiva. De esta forma, la hipótesis de estacionariedad de una serie puede ser evaluada analizando si el valor absoluto de ρ es estrictamente menor que 1. Pues bien, el test DF plantea, sencillamente, contrastar estadísticamente si $\rho = 1$.

Puesto que en economía las series explosivas no tienen mucho sentido, esta hipótesis nula se analiza frente a la alternativa que establece que:

$$H_1: \rho < 1$$

Se puede demostrar que una expresión equivalente a la anterior, mediante la toma de diferencias, viene dada por:

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

Donde $\gamma = \rho - 1$ y las hipótesis nula y alternativa son: $H_0: \gamma = 0$

$$H_1: \gamma < 0$$

Aunque el test calcula un estadístico t sobre el valor estimado de γ , bajo la hipótesis nula de existencia de una raíz unitaria, este estadístico no tiene

la distribución convencional de la t-Student. Por ello, Dickey y Fuller simularon los valores críticos para una selección de distintos tamaños muestrales, simulación ampliada posteriormente por MacKinnon.

Por lo tanto, aunque el primer paso, para la detección de raíces unitarias debe ser especificar un test DF simple, y después contrastar estadísticamente las características de ruido blanco de los residuos (estadístico Q de Ljung-Box).

Para estos casos disponemos del test ampliado de Dickey-Fuller (ADF), el cual permite considerar otros esquemas de autocorrelación. Para ello, el test ADF añade términos diferenciados de la variable dependiente y en el lado derecho de la regresión:

$$\Delta y_t = \mu + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

Esta especificación aumentada del test se basa en las siguientes hipótesis nula y alternativa:

$$H_0: \gamma = 0$$

$$H_1: \gamma < 0$$

Un resultado importante obtenido por Fuller es que la distribución asintótica del estadístico t sobre γ es independiente del número de retardos de la serie en primeras diferencias incluidos en la regresión del test ADF.

El test de Phillips Perron (PP) es un método no paramétrico para controlar la correlación de orden elevado en una serie.

PREDICCIÓN DE UNA SERIE DE DIFERENCIA

Si estimamos un modelo ARIMA con un número de diferencias, entonces será preciso recuperar las predicciones de la serie original a partir de las predicciones elaboradas para la serie de diferencias. Ellos se pueden realizarse de la forma: supongamos que Y_t denota la serie en cuyo análisis estamos interesados, y que se ha especificado y estimado el modelo univariante para la serie de primeras diferencias (Peterson, 1993).

Entonces, es claro que:

$$E_t Y_{t+k} = E_t Y_{t-k} - E_t Y_{t+k+1}$$

Por lo que:

$$\begin{aligned} E_t Y_{t+k} &= E_t Y_{t-k} - E_t Y_{t+k+1} \\ &= Y_{t+k} + Y_{t+k+1} + Y_{t+k+2} + \dots + Y_{t+k+l} + e_t \end{aligned}$$

ERROR DE PREDICCIÓN

El error de predicción es la diferencia entre la realización de la variable aleatoria y la predicción hecha para dicho valor. El error cometido en la predicción de Y_{t+k} depende del periodo en que dicha predicción se realiza.

SANGRE

La sangre es un tejido conectivo líquido, que circula por capilares, venas, arterias, aurículas y ventrículos de todos los vertebrados. Su color rojo característico es debido a la presencia del pigmento hemoglobínico contenido en los eritrocitos, es un tipo de tejido conjuntivo especializado,

con una matriz coloidal líquida y una constitución compleja. Tiene una fase sólida ,que incluye a los eritrocitos (o glóbulos rojos), los leucocitos (o glóbulos blancos) y las plaquetas, y una fase líquida, representada por el plasma sanguíneo, estas fases son también llamadas componentes sanguíneas, los cuales se dividen en componente sérico (fase líquida) y componente celular (fase sólida), su función principal es la logística de distribución e integración sistémica, cuya contención en los vasos sanguíneos (espacio vascular) admite su distribución hacia prácticamente todo el organismo (MINSA, 2012).

COMPONENTES DE LA SANGRE

PLASMA: Es un líquido transparente y de color amarillento. Contiene agua (90%), vitaminas, anticuerpos, proteínas, factores de coagulación, hormonas, sales minerales, grasas disueltas, hidratos de carbono y sus funciones son varias:

Transportadora.- De células, de materias nutritivas, de residuos, de hormonas.

Reguladora.- De la temperatura, de la presión osmótica, del equilibrio iónico, etc.

GLÓBULOS ROJOS O HEMATÍES: Son las células sanguíneas más numerosas. Contienen la hemoglobina que es la responsable de su color rojo característico. Transportan el oxígeno de los pulmones a los tejidos y recogen de estos el anhídrido carbónico para que sea eliminado en los pulmones.

Cumplen, por tanto, la función respiratoria y tienen una vida media de cuatro meses. Cada persona tiene entre 4.000.000 y 6.000.000 por mm cúbico.

GLÓBULOS BLANCOS O LEUCOCITOS: Son menos numerosos que los glóbulos rojos. Son los encargados de proteger al organismo contra los diferentes agentes patógenos, función inmunitaria o defensiva, capturándolos y destruyéndolos o produciendo anticuerpos. Una persona tiene entre 4000 y 10.000 leucocitos por milímetro cúbico. En caso de infección aumentan el número para mejorar las defensas. Algunos tienen una vida muy corta, menos de 24 horas.

PLAQUETAS O TROMBOCITOS: Son las células sanguíneas más pequeñas. Intervienen cuando se produce una rotura de la pared de los vasos sanguíneos. Se adhieren rápidamente al sitio en el que hay que parar la hemorragia, dando tiempo a la formación del coágulo definitivo, función hemostática. Tenemos entre 150.000 y 400.000 por milímetro cúbico. Tienen una vida media de 4 a 5 días.

GRUPOS SANGUÍNEOS

Los grupos sanguíneos son sistemas de clasificación de los antígenos de los glóbulos rojos. *Los glóbulos rojos tiene en su superficie algunas sustancias (antígenos) que son expresión de genes heredados de la generación anterior y que su presencia significa que el individuo heredó el de uno o de ambos padres, reaccionan con reactivos específicos en el laboratorio, es necesario conocerlos porque las transfusiones deben*

realizarse, en lo posible, con componentes del mismo grupo sanguíneo (Programa Nacional de Hemoterapia y Banco de Sangre, 2014).

El sistema ABO

En este caso la sustancia que determina el grupo sanguíneo son los azúcares, y según su composición encontramos cuatro grupos: A, B, AB y O. En cada uno de estos grupos los hematíes tienen un antígeno que los diferencia, el grupo A tiene el antígeno A, el grupo B tiene el antígeno B, el grupo AB tiene los dos antígenos y el grupo O no tiene antígeno A, ni B.

El sistema Rh

En 1940 se descubrió otro grupo de antígenos (D) que se denominaron factores Rhesus (factores Rh) porque fueron descubiertos durante unos experimentos con simios del tipo *Macacus Rhesus*. Según este grupo sanguíneo, las personas con factores Rhesus en su sangre se clasificarían como Rh positivos; mientras que aquellas sin los factores se clasificarían como Rh negativos, y sólo podrán recibir sangre de donantes Rh negativos.

TIPOS DE GRUPOS SANGUINEOS.

Según la Asociación Americana de Bancos de Sangre, la distribución de tipos de sangre y la compatibilidad es la siguiente:

Tabla N° 2: Tipos de Sangre y Compatibilidad.

Grupo	A quién puede donar	De quién puede recibir
A+	Puede donar a A+ y AB+	Puede recibir de A± y O±
A-	Puede donar a A± y AB±	Puede recibir de A- y O-
B+	Puede donar a B+ y AB+	Puede recibir de B± y O±
B-	Puede donar a B± y AB±	Puede recibir de B- y O-
AB+	Puede donar a AB+	Receptor universal
AB-	Puede donar a AB±	Puede recibir de A-, B-, AB- y O-
O+	Puede donar a A+, B+, AB+ y O+	Puede recibir de O±
O-	Donante universal	Puede recibir de O-

2.3. DEFINICIÓN DE TÉRMINOS BÁSICOS

a) Banco de Sangre

La definición establecida por la Resolución Ministerial 1191- 2006/MINSA.

Es el establecimiento autorizado para obtener, recolectar, conservar, aplicar y proveer sangre humana, así como para analizar y conservar, aplicar y proveer componentes de la misma (MINSA & Banco de Sangre, 2015).

b) Correlogramas

Es una representación gráfica de los valores individuales de la función de autocorrelación total y parcial respecto a los rezagos.

c) Donación de Sangre

Cuando una persona pierde sangre en gran cantidad por un accidente o una operación, o tiene problemas de salud, puede que sea necesario que reciba una transfusión de sangre. Sin embargo, dado que la sangre humana es una sustancia que actualmente no se puede sintetizar, es necesario extraerla de otra persona, es decir, un donador de sangre. Según (Organización Mundial de Salud, 2015).

d) Enfermedad

Enfermedad, puede definirse como una alteración de estado normal de salud, es decir pérdida transitoria o permanente del bienestar físico, psíquico o social.

e) Estacionariedad

Es una serie de tiempo, decimos que la serie es estacionaria si $f(Y_t) = f(Y_{t+k})$, es decir el comportamiento de la variable en el tiempo es el mismo si se produce un desplazamiento de la serie.

f) Estacionalidad

Puede definirse como la repetición de un cierto patrón de comportamiento en forma periódica; por ejemplo, se puede repetir cada 3 meses, 6 meses, cada año, cada 4 años, etc.

g) Grupo Sanguíneo

Un grupo sanguíneo es una clasificación de la sangre de acuerdo con las características presentes o no en la superficie de los glóbulos rojos y en el suero de la sangre. Las dos clasificaciones más importantes para describir grupos sanguíneos en humanos son los antígenos (el sistema ABO) y el factor Rh.

h) Hospital

Se denomina Hospital al lugar donde se atiende a los individuos que padecen una determinada enfermedad y que acuden a él con el objetivo de

recibir un diagnóstico y posteriormente para su tratamiento para su afección.

i) Población

Se define población como la colección completa de elementos acerca de los cuales se desea conocer la información.

j) Pronóstico

Son las predicciones de los hechos y condiciones futuras se llaman pronósticos y el acto de hacer tales predicciones se denomina pronóstico.

k) Aleatorio

Un paseo aleatorio es un proceso estocástico Y_t cuyas primeras diferencias toman un proceso de ruido blanco.

l) Ruido Blanco

Es un proceso puramente aleatorio en donde las variables son distribuidas con media cero, varianza constante y ausencia de autocorrelación entre observaciones.

m) Sangre

La sangre es un tejido conectivo líquido, que circula por capilares, venas y arterias de todos los vertebrados. Su color rojo característico es debido a la presencia del pigmento hemoglobínico contenido en los eritrocitos.

n) Servicio: Nuevo .continuidad, reingreso

o) Sangre total

Sangre total pueden separarse varios componentes en el mismo banco de sangre. Los hematíes y las plaquetas se aíslan de la sangre total mediante centrifugación suave, siendo posteriormente procesados para obtener varios preparados distintos (Unidad de Hematología y Trasfusión, 2015).

p) Sangre y sus componentes hemoderivados

Entendemos por componente sanguíneo al producto separado de una unidad de sangre total, mientras que la denominación derivado del plasma hace referencia a un producto separado de un gran volumen de mezclas de plasma mediante un proceso llamado fraccionamiento (Hemoderivados, 06 de Julio 2015).

q) Salud

Es definida por la constitución de 1946 de la Organización Mundial de Salud como, el estado de completo bienestar físico, mental y social.

r) Serie

Una serie de tiempo es un conjunto de observaciones ordenadas en el tiempo (o en alguna otra dimensión).

s) Transfusión de Sangre

Una transfusión de sangre es la transferencia de sangre o componentes sanguíneos de un sujeto (donante) a otro (receptor). Una transfusión de sangre puede salvar la vida del paciente, de ahí la necesidad de que los servicios de salud procuren mantener un suministro adecuado de sangre segura y garantizar que se utilice como corresponde.

t) Unidad de Sangre

El plasma procedente de una unidad de sangre total, aislado y congelado con la suficiente rapidez después de la donación, como para que se mantenga la actividad de los diferentes factores de la coagulación, aproximadamente por encima del 70 % de la inicial (OMS, Donación de Sangre & Banco de Sangre, 2015).

u) Variable Dependiente

Son variables que influyen en el conjunto de relaciones y a su vez están influenciados por las variables independientes.

v) Variable Independiente

Variabes que influyen en el conjunto de relaciones pero no están influenciados por ella.

2.4. OPERACIONALIZACIÓN DE VARIABLES

Tabla N° 3: Operacionalización de Variables.

Variables	Indicador	Unidad de medida
Variable Dependiente: Volumen del número de unidades de transfusión de sangre.	Unidades de transfusión de sangre.	Unidad de sangre (ml)
Variabes Independiente: Volumen retardado.	Unidades de transfusión de sangre	Unidades de sangre (ml)

Fuente: Elaboración propia.

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1. LOCALIZACIÓN

El presente trabajo de investigación se realizó en el Hospital Regional Manuel Núñez Butrón -Puno, específicamente en el centro de Hemoterapia y banco de sangre.

3.2. MATERIALES Y MÉTODOS

Para el presente trabajo de investigación se obtuvo la información de la Unidad de Estadística e Informática en el Hospital Regional Manuel Núñez Butrón –Puno, periodo 2006 - 2015-I.

Se usó equipos de cómputo como: laptop Core i5, impresora Epson L355 y para el cálculo y procedimientos de datos google Drive más conocido como “cloud computing” y software libre Gretl 1.10.1.

3.3. POBLACIÓN

La población para el siguiente estudio estará constituida por el número total de pacientes registrados mensualmente en el libro de informe de centro de hemoterapia y banco de sangre en el Hospital Regional Manuel Núñez Butrón Puno.

3.4. MUESTRA

La muestra estará conformada por 144 pacientes registrados mensualmente en el libro de informe de Centro de Hemoterapia y Banco de Sangre en el Hospital Regional Manuel Núñez Butrón Puno en el periodo 2006- 2015-I.

3.5. UNIDAD MUESTRAL

Número de unidades de transfusión de sangre

3.6. MÉTODOS DE RECOPIACIÓN DE DATOS

Con la finalidad de asegurar la efectividad y originalidad de los datos, estos datos fueron recopilados directamente del registro de relación de pacientes que recibieron transfusión de sangre, que se encuentran registrados diariamente en el libro de Centro de Hemoterapia y Banco de Sangre en el Hospital Regional Manuel Núñez Butrón Puno, 2006-2015-I.

3.7. METODOLOGÍA

El presente trabajo de investigación es analítico- aplicativo; este método es de gran importancia, ya que está presente en todas las fases de desarrollo de los modelos.

METODOLOGÍA DE ANÁLISIS DE DATOS

Para el presente trabajo de investigación se hizo uso de la metodología de Box-Jenkins, más conocido como enfoque de Box-Jenkins en las series de

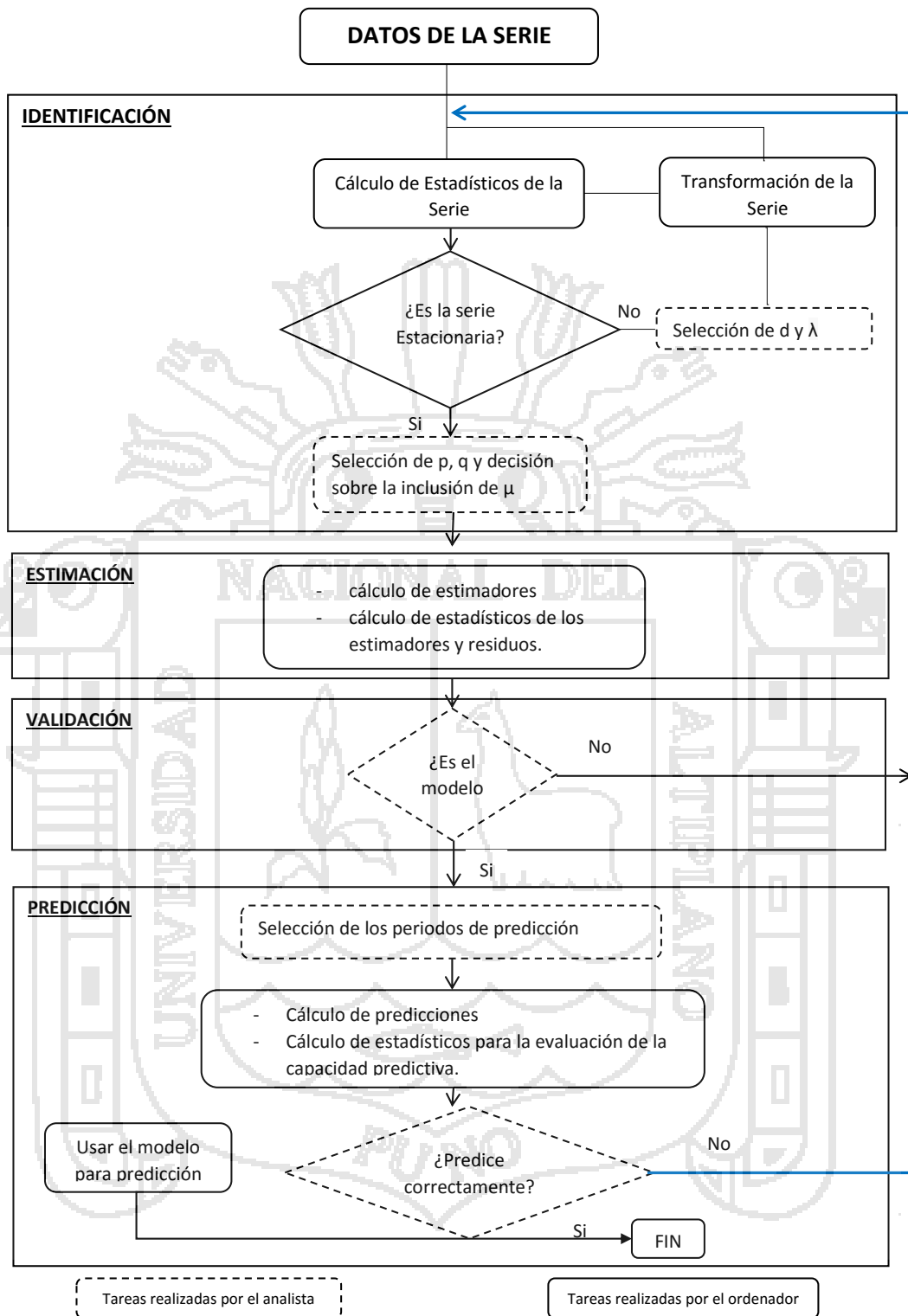
tiempo de los denominados modelos ARIMA, los pasos a seguir en la obtención del modelo univariante por el método de Box-Jenkins fueron:

- 1) Representación gráfica de la serie
- 2) Histograma
- 3) Normalidad de datos transformados
- 4) Calculo de la función de autocorrelacion (F.A.C.) y la función de autocorrelacion parcial (F.A.C.P.)
- 5) Proceso de identificación
- 6) Estimación de parámetros
- 7) Proceso de verificación
- 8) Proceso de predicción

METODOLOGÍA DE BOX-JENKINS (teoría de WIENER-KOLMOGOROV)

Para el modelo de ajuste de unidades de transfusión de sangre en el Hospital Regional Manuel Núñez butrón Puno, se aplicó la metodología Box-Jenkins (teoría de WIENER-KOLMOGOROV) sigue un proceso que consta de cuatro fases.

Figura N° 2 : Metodología del Enfoque Box- Jenkins



Fuente: (Uriel, 1985) Análisis de series temporales Modelos ARIMA, (Pag.72)

FASES DE LA METODOLOGÍA BOX-JENKINS

IDENTIFICACIÓN DEL MODELO

El primer paso en la identificación del modelo consiste en determinar si la serie es estacionaria en función de la serie original. Si la serie no es estacionaria, se convierte a una serie estacionaria. Se estabiliza la serie mediante la transformación de Box Cox, se especifica el grado de diferenciación según al número de veces que la serie sea diferenciada para obtener una serie estacionaria en media y varianza. El orden de integración define el parámetro "d" del modelo ARIMA. En la práctica es suficiente tomar una diferencia (d=1), o dos diferencias (d=2) para obtener una serie estacionaria en media, A partir de la tercera diferencia la varianza se deforma, es decir la varianza crece. Un aspecto importante en la modelación ARIMA de una serie de tiempo simple es el número de veces que necesita de una diferencia antes de fijar el modelo. Para esto se utiliza la prueba de Dickey Fuller o de raíces unitarias en la cual la hipótesis a verificar es:

H_0 = Hay raíz unitaria (proceso no estacionario).

H_1 = No hay raíz unitaria (proceso estacionario).

EL estadístico de prueba es: $T_u = a/\hat{s}(\hat{a}_0)$ Donde:

a_0 : Estimación mínimo cuadrado de a. La regla de decisión a considerar

Será:

Rechazar H_0 si $T_u > T_1$ y por lo tanto el proceso es estacionario. El estadístico T_u sigue una distribución T de Student que se tabula por Dickey y Fuller.

Una vez obtenida una serie estacionaria con algún valor para la diferencia “d”, se deberá de identificar la forma del modelo a utilizar encontrando los valores apropiados de p y q, mediante la comparación de los coeficientes de autocorrelacion parcial de los datos. Si el modelo no es estacionario, se puede intentar un modelo alternativo.

ESTIMACIÓN DE PARAMETROS

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 de media móvil incluidos en el modelo, a través del uso de los mínimos cuadrados no lineales que minimiza la suma de los cuadrados de los residuos.

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 blando.
- b) El modelo estimado es estacionario e invertible.
- c) Los coeficientes son estadísticamente significativos.

d) Una de las formas para detectar anomalías a los supuestos, es a través del análisis de los residuales.

REALIZACION DE PRONÓSTICO CON EL MODELO

Una vez identificado y/o encontrado el modelo adecuado se puede realizar predicciones para varios periodos futuros de tiempo.

Estimados los parámetros del modelo encontrado correspondiente y después que haber pasado la etapa de verificación se utiliza el modelo estimado en la predicción de valores futuros de la variable objeto de estudio.

Si la serie cambia a través de tiempo como pudiera ser necesario de calcular los parámetros o incluso desarrollar un nuevo modelo.

PROCESO DE PREDICCIÓN

Dada una serie estacionaria que sigue cualquier proceso de la forma:

$$z_t = \phi_1 z_{t-1} + \dots + \phi_p z_{t-p} + E_t + \theta_1 E_{t-1} + \dots + \theta_q E_{t-q}$$

Si se tiene la información hasta el momento t y para predecir "m" periodos hacia delante, se construye la función.

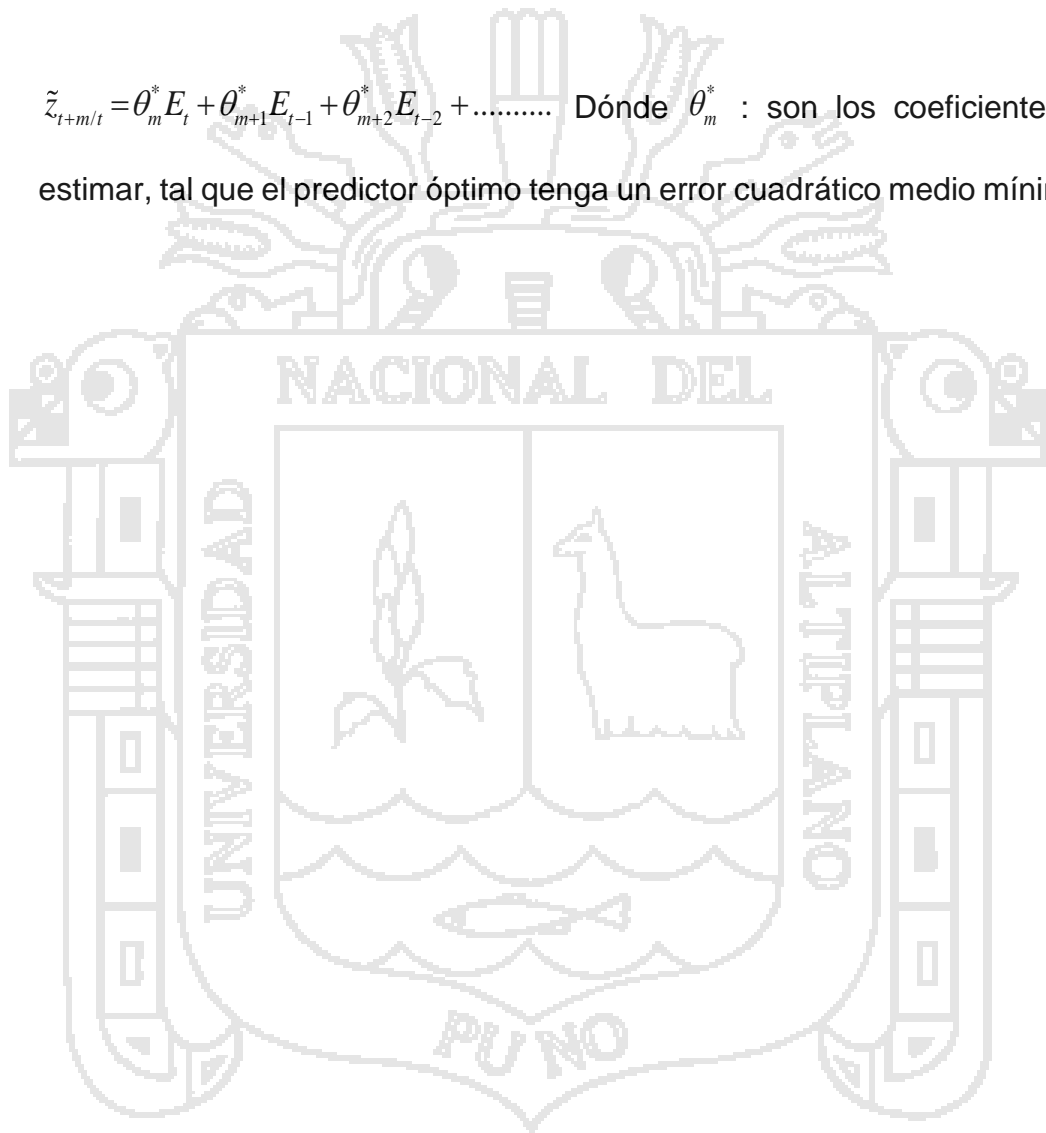
$$z_{t+m} = \phi_1 z_{t+m-1} + \dots + \phi_p z_{t+m-p} + E_{t+m} + \theta_1 E_{t+m-1} + \dots + \theta_q E_{t+m-q}$$

Cuya función eficiente de predicción sea $z_{t+m/t}$ cumple con los siguientes supuestos.

- ✓ Los parámetros ϕ_p y θ_q son conocidos.
- ✓ Los errores pasados y presentes son conocidos.

$E_t, E_{t-1}, \dots, E_1, E_1$ Predictor óptimo se construye de la función lineal de todos los valores conocidos de E_t ,

$\tilde{z}_{t+m/t} = \theta_m^* E_t + \theta_{m+1}^* E_{t-1} + \theta_{m+2}^* E_{t-2} + \dots$ Dónde θ_m^* : son los coeficientes a estimar, tal que el predictor óptimo tenga un error cuadrático medio mínimo.



CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

Para el análisis de datos, se ha utilizado la información del número de unidades de transfusión de sangre del periodo 2006 – 2015-I, los cuales fueron obtenidos de la Oficina de Estadística e Informática del Hospital Regional Manuel Núñez Butrón- Puno. Se muestra los datos originales del número de unidades de transfusión de sangre del periodo 2006 – 2015-I, en anexo 1.

Para un mejor tratamiento de información

- ☞ Ordenamiento de información.
- ☞ Graficar la serie para explorar los movimientos que tiene.

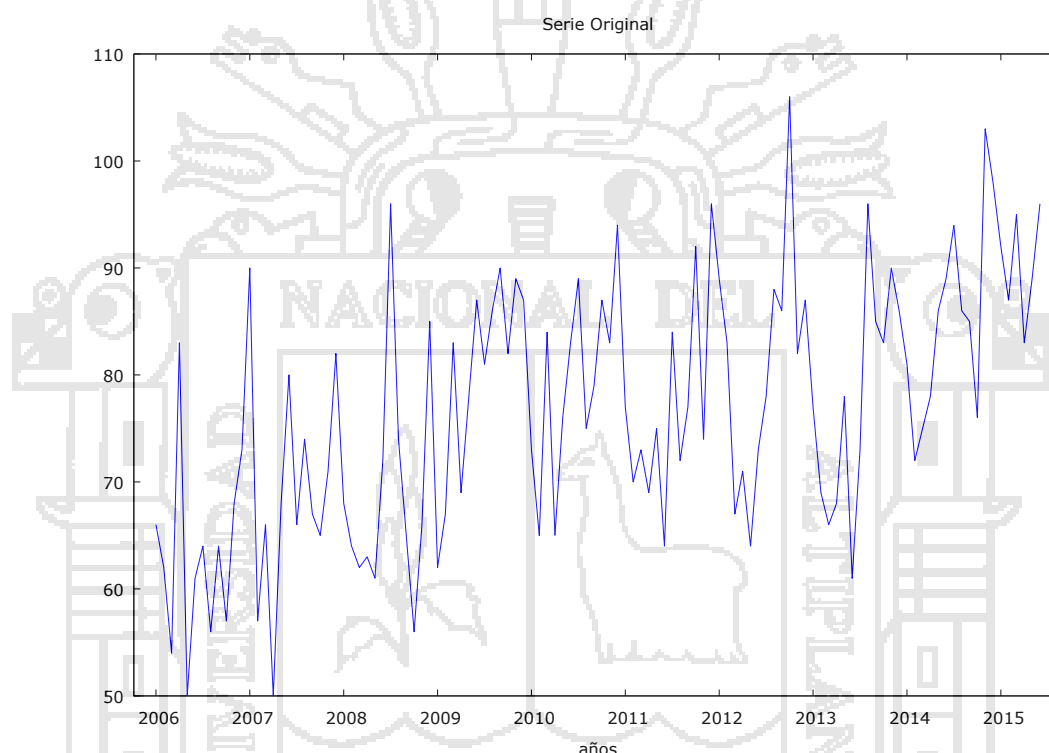
Análisis e interpretación

Se realiza el modelo univariante de número de unidades de transfusión de sangre en el Hospital Regional Manuel Núñez Butrón Puno y sus respectivas variables explicativas mediante el uso de software libre, Grelt 1.10.1.

4.1. ANALISIS ESTADÍSTICO DE LA SERIE TIEMPO (ESTUDIO EXPLORATORIO)

4.1.1. FASE DE IDENTIFICACIÓN DEL MODELO

Gráfico N° 1: número de unidades de transfusiones de *sangre*,
periodo 2006-2015-I.



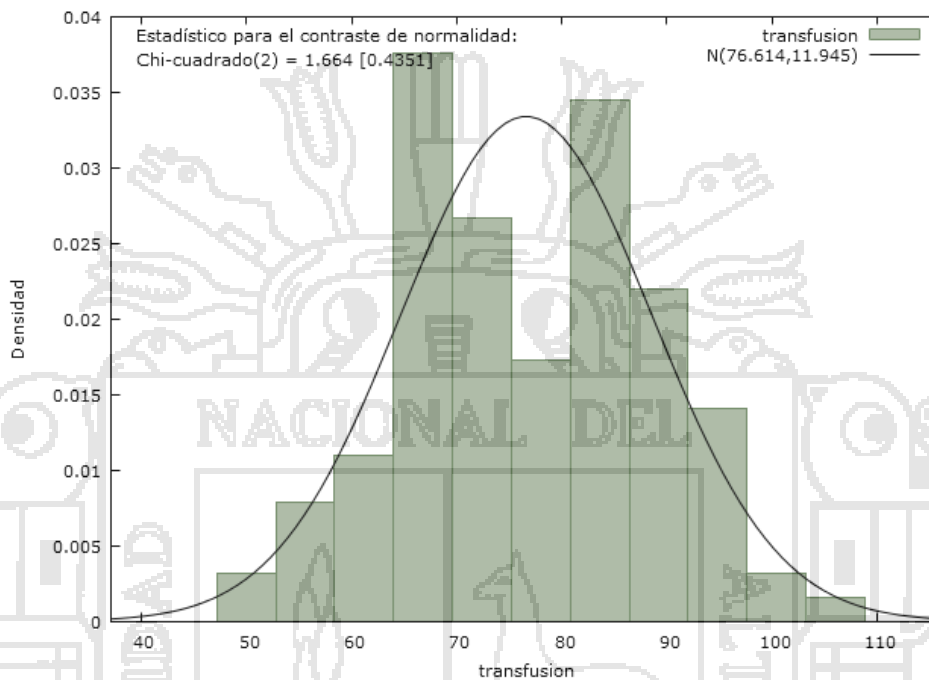
Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

El gráfico N° 1, Evolución de la serie del número de unidades de transfusiones de sangre del Hospital Regional Manuel Núñez Butrón Puno, 2006 -2015-I, ha sido obtenido del software libre Gretl versión 1.10.1.

El modelo univariante de la serie de número de unidades de transfusiones de sangre en el Hospital Regional Manuel Núñez Butrón Puno, 2006- 2015-I, del gráfico N°1, podemos afirmar que la serie muestra muchos picos, los

cuales parecen estar espaciados uniformemente con una tendencia lineal creciente, sugiriendo la presencia de un componente periódico de la serie.

Gráfico N° 2: Histograma de Serie de tiempo del número de Unidades de transfusiones de sangre.



El gráfico N° 2, se observa la histograma de la serie del número de unidades de transfusiones de sangre del Hospital Regional Manuel Núñez Butrón Puno, 2006 -2015-I, ha sido obtenido del software libre Gretl versión 1.10.1.

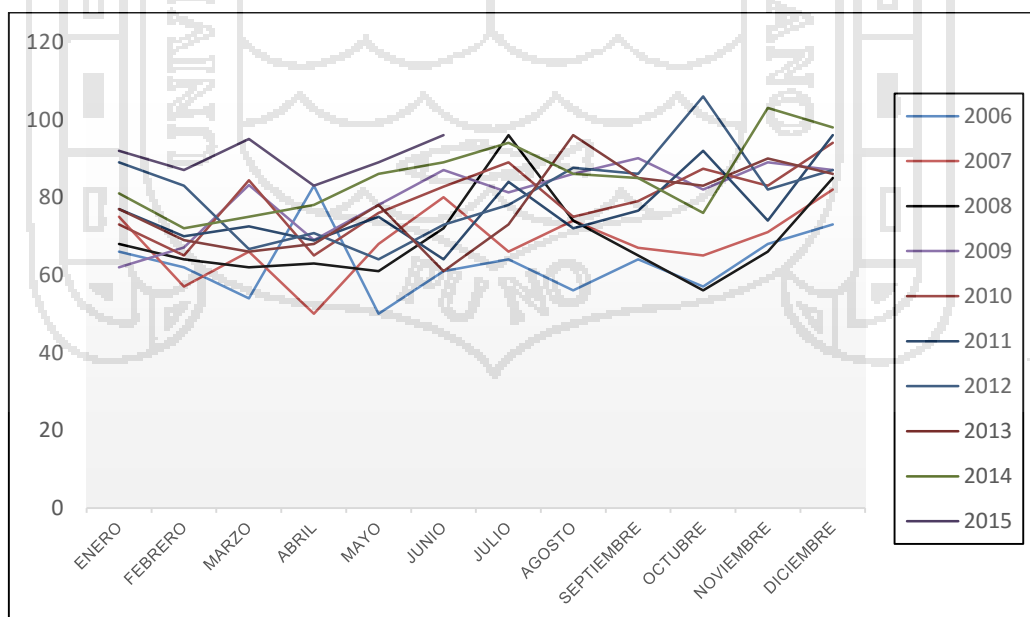
El gráfico N° 2, se observa la histograma de la serie del número de unidades de transfusiones de sangre no se distribuye según una distribución normal.

Y podemos afirmar que los valores individuales de la serie no giran en torno a su valor medio, por lo tanto la serie no es estacionaria esto nos indica que se debe de realizar la transformación de la serie.

Mientras que (Cutipa, 2010) indica que la serie de atenciones por emergencia de Cirugía General no es estacionaria en el nivel puesto que los valores se distribuyen debajo de la media, también la serie representa una no estacionariedad de manera que se produce un crecimiento o decrecimiento en algunos periodos, esto indica que se debe de realizar la transformación de la serie.

Como ambas son similares, decir los pacientes ingresantes al servicio de Cirugía general, sea por operaciones de accidentes de tránsito, apéndices, aborto, etc. Además coincidimos en la tendencia que presenta esta serie, por el crecimiento de la población.

Gráfico N° 3: Serie del número de transfusiones de sangre por años



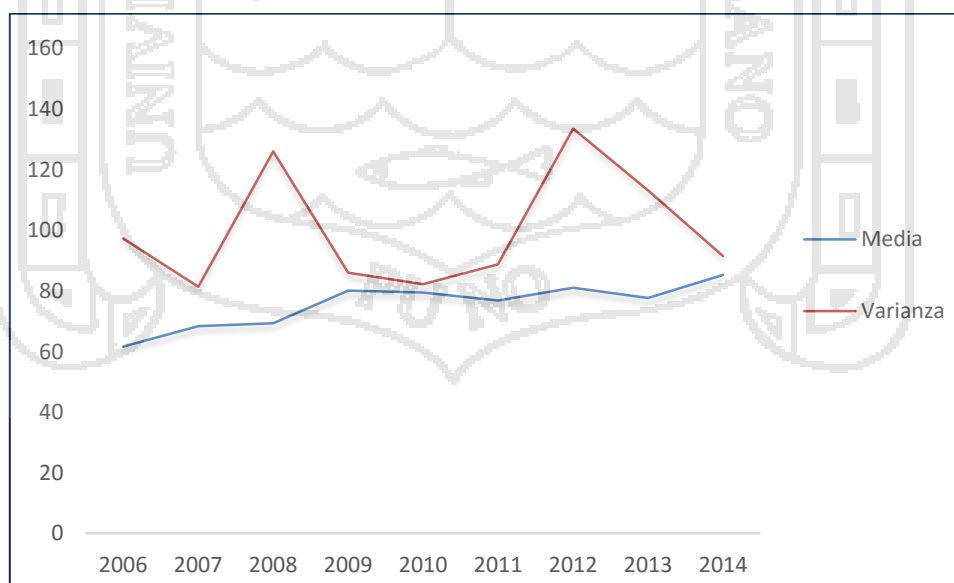
Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

En el gráfico N°3, serie de número de unidades de transfusión de sangre por años obtenidas de Google Drive.

Observamos que la serie tiene valores individuales, es decir no giran en torno a su nivel medio y aparentemente la serie de número de unidades de transfusión de sangre tiene máximos y mínimos que se repiten con frecuencia para cada año, se estaría hablando de una serie estacional.

Mientras que (Centeno, 2013) la serie de Atendidos del Hospital de Apoyo de Sandía no tiene máximos ni mínimos que se repiten con frecuencia para cada año, mientras que en la serie de número de unidades de transfusión de sangre, tiene un crecimiento respecto al año anterior, considero que es debido a crecimiento de la población y la capacidad de pagar cuanto sea por recibir una atención resolutive de calidad.

Gráfico N° 4: Media y Varianza del número de unidades de transfusiones de sangre por años.



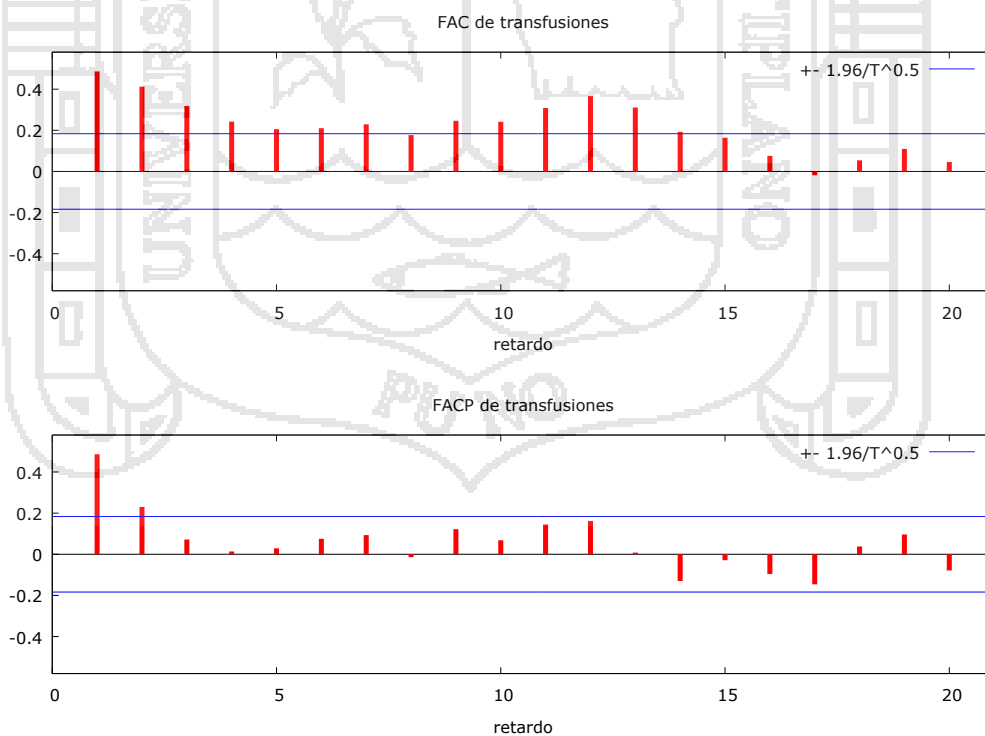
Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

En el gráfico N° 4, Media y varianza de la Serie de número de unidades de transfusión de sangre por años obtenidas de Google Drive.

En el gráfico N°4, se observa que no es estacionaria en media por la tendencia lineal creciente que presenta y no es estacionaria en varianza, por lo tanto se necesita realizar transformaciones.

Mientras que (Valdez, 2015) indica que la serie intervenciones quirúrgicas por aborto no es estacionaria en media por la tendencia positiva y es estacionaria en varianza, coincidimos en la tendencia que presenta esta serie, por el crecimiento de la población y el lugar de trabajo de investigación.

Gráfico N° 5 : Correlogramas de FAC y FACP de la serie del número de transfusiones de sangre.



Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01).

En el gráfico N°5, se obtienen las funciones de autocorrelación y autocorrelación parcial estimadas obtenidas de software libre Gretl 1.10.1.

En el gráfico N°5, podemos observar que la función de autocorrelación muestra altos en los retardos múltiplos del periodo estacional 1,12. También, se observa que las funciones de autocorrelación y autocorrelación parcial estimadas también validan los periodos estacionales, porque los coeficientes de autocorrelación para retardos múltiplos del periodo estacional de la serie son significativamente distintos de cero.

Los coeficientes de autocorrelación decaen rápidamente después de rezago 12 del periodo estacional, entonces hay un indicio de que hay estacionariedad en media.

La estructura ARIMA estacional (modelo SARIMA) puede representar alguna combinación de componentes de integración, Autorregresivos y de media móvil.

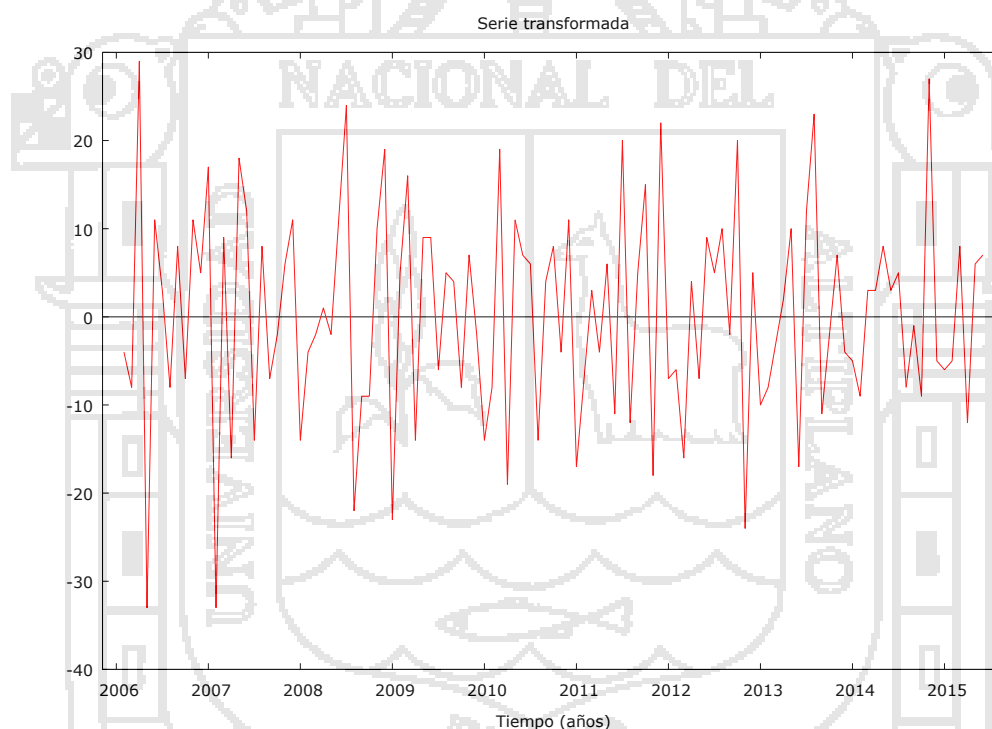
De acuerdo a estudios confirmatorios, con base al test estadístico de Box-Piere-Ljung, hay evidencia suficiente como para abonar la conclusión general de que la serie del número de unidades de transfusión de sangre en el Hospital Regional Manuel Nuñez Butrón Puno, periodo 2006 – 2015- I es no estacionaria en varianza.

Para probar la hipótesis conjunta de que todos los coeficientes de autocorrelación ρ_k son simultáneamente iguales a cero, se utilizó el estadístico de Box-Piere-Ljung.

TRANSFORMACIÓN DE LA SERIE

Para que la serie del número de unidades de transfusión de sangre del Hospital Regional Manuel Núñez Butrón Puno, periodo 2006 – 2015-I, sea estacionaria en media se debe de transformar con la primera diferencia regular a la serie del número de unidades de transfusión de sangre y tenemos como resultado lo siguiente:

Gráfico N° 6 : Primera diferencia regular de la serie del número de unidades de transfusión de sangre.



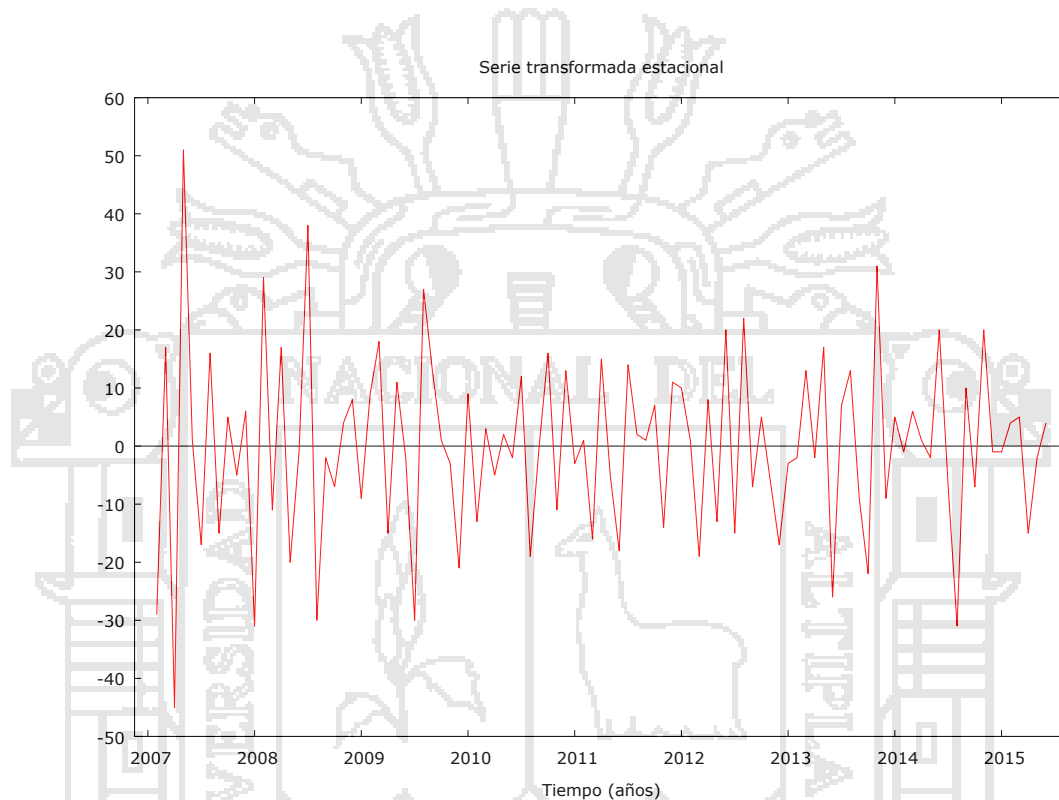
Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

En el gráfico N° 6, la primera diferencia regular de la serie del número de unidades de transfusión de sangre, obtenida de software libre Gretl 1.10.1.

Observando el gráfico N°6, después de realizar la primera diferencia regular aun presenta variaciones altas y bajas, presenta picos repetitivos

de periodo $s=12$ es decir tiene una variación alta en el mes de Diciembre del 2006 como también como también en mes de Diciembre del 2008, por lo tanto se realizó otra diferencia adicional pero en la parte estacional.

Gráfico N° 7: Primera diferencia regular y estacional de la serie del número de unidades de transfusión de sangre.



Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

En el gráfico N° 7, la primera diferencia regular y primera diferencia estacional de la serie del número de unidades de transfusión de sangre, obtenida de software libre Gretl 1.10.1.

Después de transformaciones de la serie se puede observar que la volatilidad de los datos es mínimo aunque presenta variabilidades altas y bajas en las observaciones 16, 32,95 y 104. Realizando nuevamente la prueba de estacionariedad en media y varianza según contraste de la

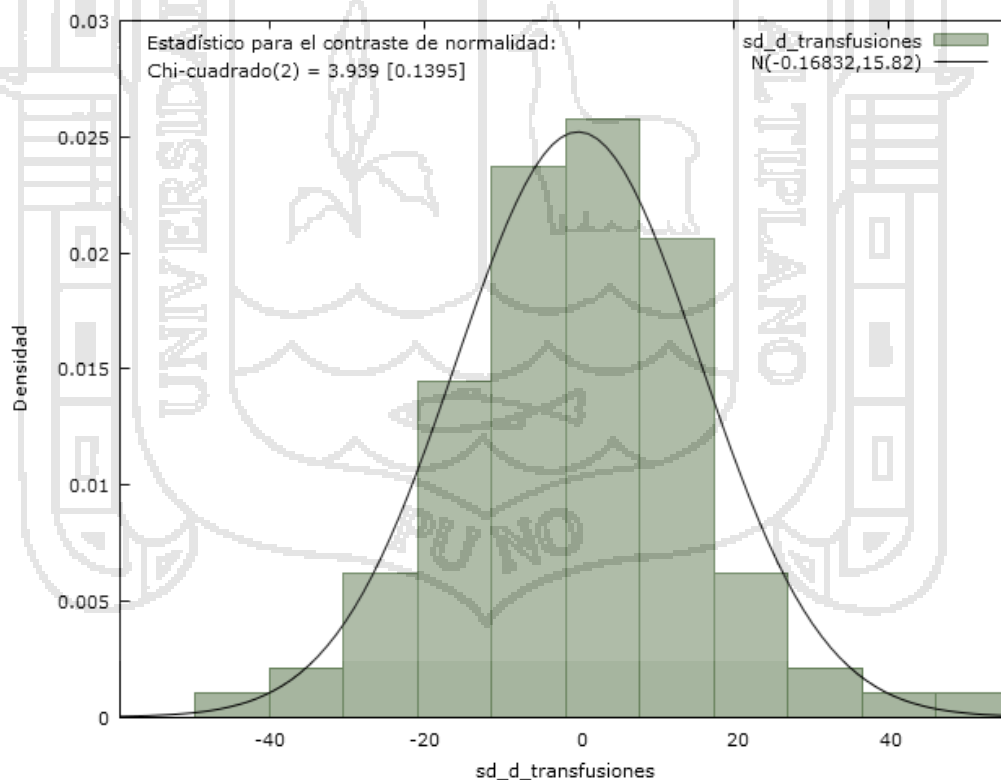
hipótesis nula de distribución normal: Chi-cuadrado (2) = 3.939 con valor p 0.13949 por lo tanto la serie es estacionaria en media y varianza.

Prueba de normalidad de la serie del número de unidades de transfusión sangre transformadas.

H_0 : La serie del número de unidades de transfusión de sangre se distribuyen según una distribución normal.

H_1 : La serie del número de unidades de transfusión de sangre no se distribuyen según una distribución normal.

Gráfico N° 8: Normalidad de la serie del número de unidades de transfusión de sangre transformadas.

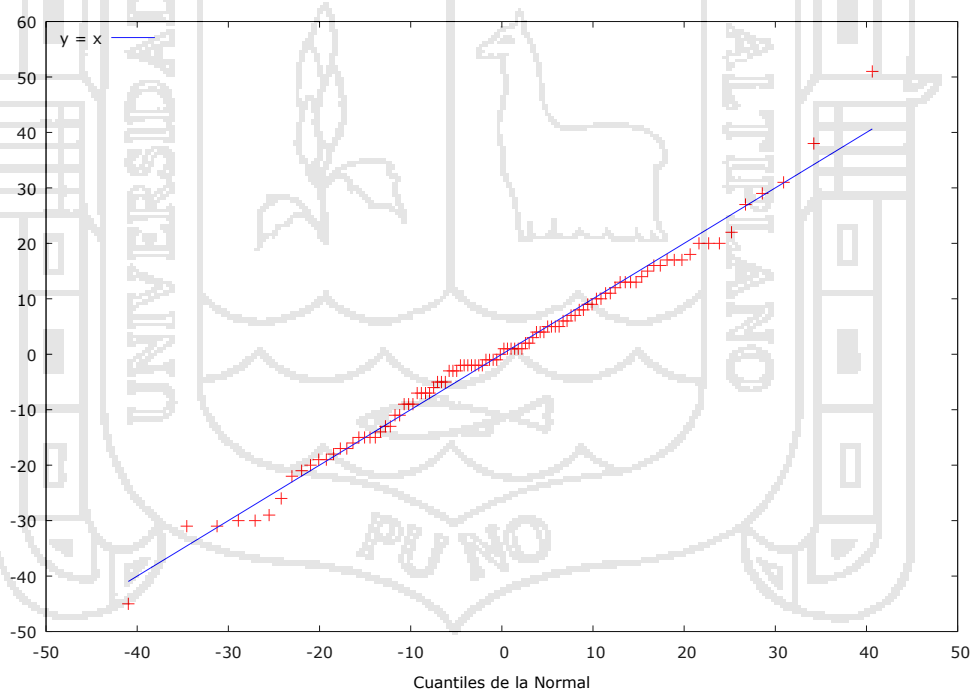


Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

En el grafico N° 8, de la normalidad se la serie del número de unidades de transfusión de sangre, ha sido obtenida de software libre Gretl 1.10.1.

Observando el grafico N° 8, el valor de $P > 0.05$, como el valor de P es mayor se acepta la hipótesis nula, entonces confirmamos que la serie del número de unidades de transfusión de sangre se distribuyen según una distribución normal y confirmamos que los valores individuales de la serie giran en torno a su valor media, entonces la serie del número de unidades de transfusión de sangre en el Hospital Regional Manuel Núñez Butrón Puno, periodo 2006- 20015-I, es estacionaria.

Gráfico N° 9: Q-Qplot Cuartiles normales de la serie del número de unidades de transfusión de sangre transformadas.



Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

En el grafico N° 9, Cuartiles normales se la serie del número de unidades de transfusión de sangre, ha sido obtenida de software libre Gretl 1.10.1.

De acuerdo a estudios confirmatorios de la prueba de Dickey Fuller y prueba de Phillips Perron para la serie es estacionaria y procedimiento de estadístico pierce de Q' Ljung-Box para estacionalidad.

PRUEBA ANALÍTICA DE ESTACIONALIDAD DICKEY FULLER

$H_0: \rho=1$ La serie tiene raíz unitaria (Serie no Estacionaria).

$H_a: \rho < 1$ La serie no tiene raíz unitaria (Serie Estacionaria).

El reporte de Dickey Fuller en Gretl 1.10.1 nos muestra lo siguiente:

Tabla N° 4: Prueba de raíz unitaria para la serie del número de unidades de transfusión de sangre.

Prueba de raíz unitaria	prueba de Dickey Fuller	Prueba de Phillips - perron
H_0 : Serie no Estacionaria	a=1	
contraste con constante		
Modelo:	$(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$	
Estadístico de Contraste:	-3.86883	-6.90108
Valor p asintótico:	0.002288	3.122e-011

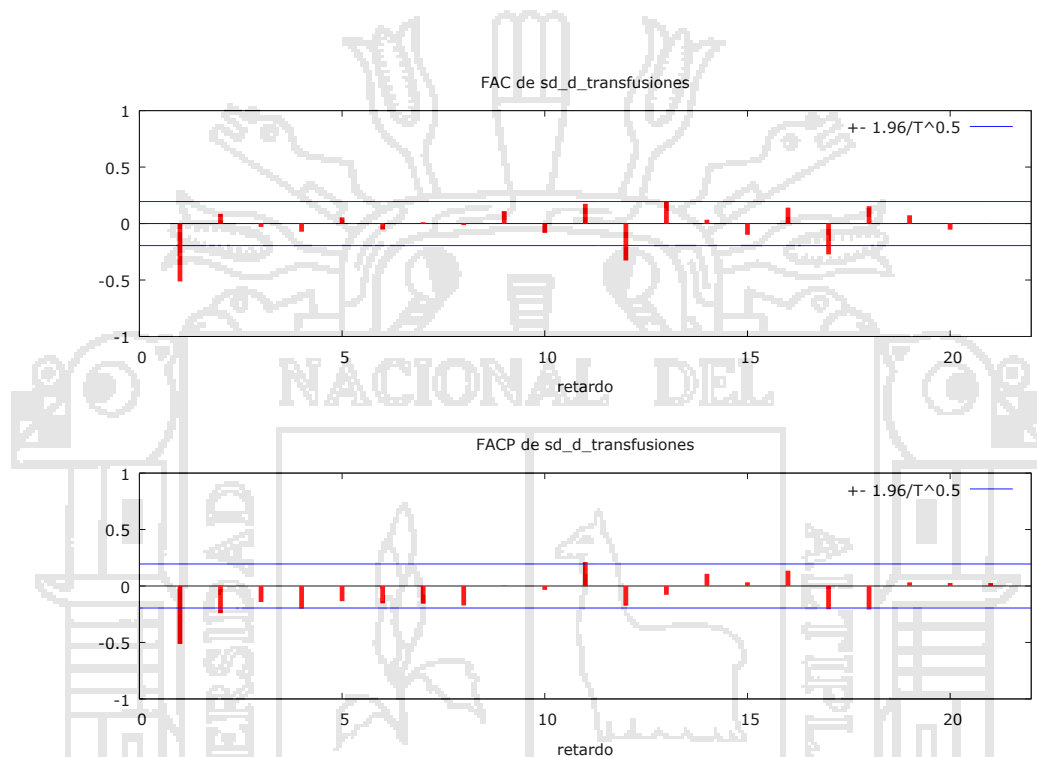
Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

En la tabla N°4, Prueba de raíz unitaria para serie del número de unidades de transfusión de sangre, ha sido obtenida de software libre Gretl 1.10.1.

Como $p = 0.002$ es menor que 0.01 y 0.05, se acepta la hipótesis alterna, por lo tanto se confirma que la serie transformada es estacionaria. Para determinar el orden de comportamiento autorregresivo (p), así como también el orden de componente de medias móviles (q).

Para ello se procedió a observar los correlogramas de las funciones de autocorrelacion y autocorrelacion parcial con el fin de determinar el modelo óptimo de proyección para el número de unidades de transfusión de sangre.

Gráfico N° 10: Correlogramas de FAC y FACP de la serie del número de unidades de transfusión de sangre transformadas.



Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

En el gráfico N° 10, de la función de autocorrelacion y autocorrelacion parcial estimada para la serie del número de unidades de transfusión de sangre del Hospital Regional Manuel Núñez Butrón Puno, 2006 – 2005-1, ha sido obtenido del software libre Gretl 1.10.1.

Se observa en el gráfico N° 10, la función de autocorrelacion y autocorrelacion parcial los retardos más significativos son el primero y doce, y que las autocorrelaciones para los primeros retardos de tiempo

caen hacia cero de manera muy semejante a la de patrón teórico de un proceso AR (1) Y AR (2).

(Cutipa,2010) una vez realizada la primera diferencia regular, se llega a que los datos son estacionarios en media y varianza de la serie de atenciones por emergencia de cirugía general, esta serie al realizar la primera diferencia regular es estacionaria y mientras en la serie de número de unidades de transfusión de sangre se realiza una diferencia más en la parte estacional, para conseguir la estacionariedad en la serie, esta diferencia se debe a la toma de muestra de pocos años para el análisis de series de tiempo.

IDENTIFICACIÓN DE LA FORMA DEL MODELO A UTILIZAR

En la fase de identificación de modelo se procede a identificar los valores de la parte autorregresivo (p) y de la parte de medias móviles (q) y parte estacional de un modelo SARIMA (p, d, q) (P, D, Q)¹².

Conforme se observa en el gráfico de la parte regular se obtiene (p, d, q). La parte AR (1) proviene del decrecimiento rápido inicial de la función de autocorrelacion parcial. Así mismo la parte MA (1) proviene que la función de autocorrelacion presenta 1 retardo significativo en la mayoría de periodos.

Con sus respectivos correlogramas, para las autocorrelaciones simples y parciales, que genera este modelo, en los que se puede apreciar su comportamiento.

Los modelos sugeridos para esta serie son:

- ✓ SARIMA(0,1,1)(1,0,0)₁₂

Cuya ecuación es respectivamente:

$$\Leftrightarrow (Y_t - Y_{t-1}) - (\varphi_{12}Y_{t-12} - \varphi_{13}Y_{t-13}) = (\varepsilon_t - \theta_1\varepsilon_{t-1})$$

$$\hat{Y}_t = Y_{t-1} + \varphi_{12}Y_{t-12} + \varphi_{13}Y_{t-13} + \varepsilon_t - \theta_1\varepsilon_{t-1}$$

- ✓ SARIMA (0, 1,1) (1, 1,0)₁₂

Cuya ecuación es respectivamente:

$$\Leftrightarrow (Y_t - Y_{t-1}) - (\varphi_{12}Y_{t-12} - \varphi_{13}Y_{t-13}) = (\varepsilon_t - \theta_1\varepsilon_{t-1}) -$$

$$(\varepsilon_{t-12} - \varepsilon_{t-13})$$

$$\hat{Y}_t = Y_{t-1} + Y_{t-2} + \varphi_{12}Y_{t-12} + \varphi_{13}Y_{t-13} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}$$

4.1.2.FASE DE ESTIMACIÓN DEL MODELO

En la fase de estimación de modelos, continuamos con la metodología Box-Jenkins el paso a seguir es la especificación del modelo, en el mismo que se analizó la significancia tanto individual como global de sus parámetros, esto lo hace el ordenador con el software libre Gretl 1.10.1.

Tabla N° 5: Comparativo de modelos estimados de la serie del número de unidades de transfusión de sangre.

Modelo: SARIMA(0,1,1)(1,0,0)12				Modelo: SARIMA(0,1,1)(1,1,0)12			
<i>Modelo general: $\phi_p(L)\phi_{PS}(L)\Delta^d\Delta_S^D Y_t = \theta_q(L)\omega_{QS}(L)\varepsilon_t + \delta$</i>							
$\hat{Y}_t = Y_{t-1} + 0.32Y_{t-12} + 0.32Y_{t-13} + \varepsilon_t + 0.86\varepsilon_{t-1}$				$\hat{Y}_t = Y_{t-1} + Y_{t-2} - 0.36Y_{t-12} - 0.36Y_{t-13} + \varepsilon_t + 0.89\varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}$			
Coeficiente	Z	Valor p		Coeficiente	Z	Valor p	
Phi_1	0.320745	3.480	0.0005 ***	Phi_1	-0.3618	-3.573	0.0004 ***
theta_1	-0.86312	-15.89	7.41e-057 ***	theta_1	-0.8948	-14.94	1.70e-050 ***
Criterio de Akaike	840.4587			Criterio de Akaike	781.9898		
Criterio de Schwarz	848.6408			Criterio de Schwarz	789.8352		

Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

En tabla N° 5, comparativo de modelos estimados de la serie del número de unidades de transfusión de sangre ha sido obtenido de software libre Gretl 1.10.1.

Según la hipótesis de:

$$H_0: \phi_1 = 0 \quad P > 0.05$$

$$H_1: \phi_1 \neq 0 \quad P < 0.05$$

Para el coeficiente de Phi (1), p= 0.0005 es menor que 0.05, se acepta la hipótesis alterna, por lo tanto el coeficiente es diferente de cero, por lo cual el modelo es adecuado para el pronóstico.

También se tiene que para el segundo modelo $p=0.0004$ menor que 0.05. Entonces se acepta la hipótesis alterna, por lo que el modelo también es adecuado para el pronóstico, entonces los dos modelos son adecuados para el pronóstico.

Para la selección del mejor modelo de estimación se ha computado para cada una de las series consideradas los criterios de información, por cada criterio se utilizó aquel modelo que minimiza el criterio de información el de mejor ajuste.

Como se observa en la tabla N° 5, los criterios de información Akaike y Schwarz disminuyen de 840.45 a 781.98 y de 848.64 a 789.83 respectivamente, aunque los errores tienen comportamiento ruido blanco de los modelos.

Por lo tanto el que mejor describe a la serie analizada es el modelo SARIMA $(0, 1, 1) (1, 1, 0)_{12}$. Siendo cuya ecuación: $\hat{Y}_t = Y_{t-1} + Y_{t-2} + \varphi_{12}Y_{t-12} + \varphi_{13}Y_{t-13} + \varepsilon_t - \theta_1\varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}$

4.1.3. FASE DE EVALUACIÓN O VALIDACIÓN DEL MODELO

En esta etapa de evaluación del modelo vamos a ver la adecuación entre el modelo y los datos, o dicho de otra forma veremos en qué medida los residuos del modelo estimado se aproximan al comportamiento de un ruido blanco. Como se sabe, la función de autocorrelación 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 Pierce es más favorable al ruido blanco cuanto mayor sea la probabilidad p valor. Si se utiliza al nivel de confianza habitual del 95%, entonces los residuos son ruido blanco siempre que el p- valor sea superior a 0.05, para el coeficiente de autocorrelacion de que los residuos forman una secuencia aleatoria, se realizó las siguientes comparaciones.

$H_0: p > \alpha$ Los residuales son ruido blanco.

$H_1: p < \alpha$ Los residuales no son ruido blanco.

Tabla N° 6: Prueba de Ljung-Box Q' de la serie del número de transfusiones de sangre.

Retardos	Ljung-Box Q'	Valor p
12	8.5143	0.579
24	24.3342	0.330
36	40.2438	0.213
48	60.3046	0.077

Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

En tabla N° 6, Prueba Ljung-Box Q' de la serie del número de unidades de transfusión de sangre ha sido obtenido de software libre Gretl 1.10.1.

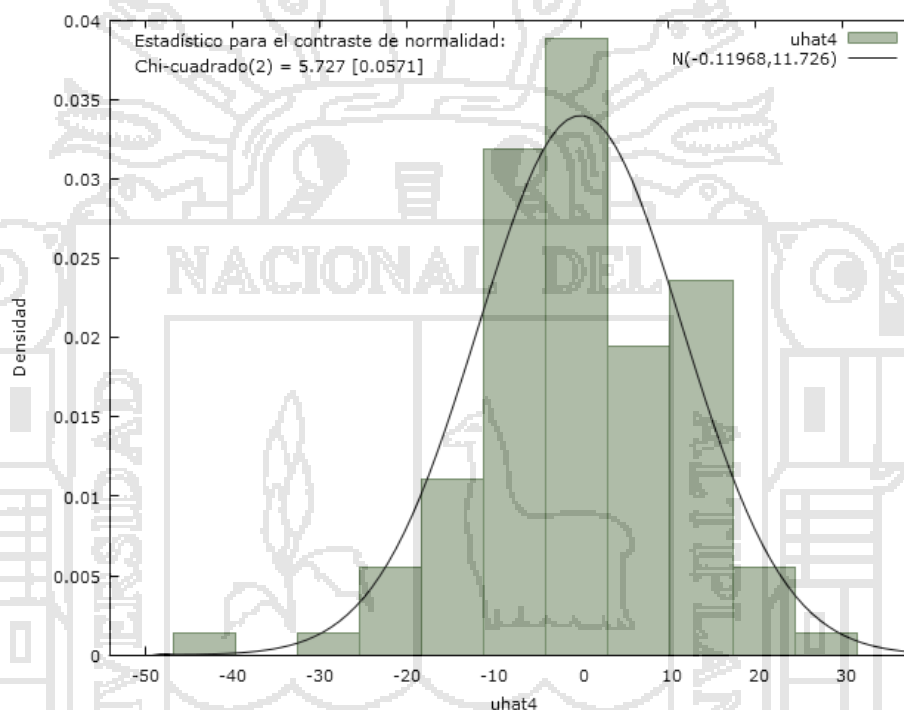
Estadístico Q de Ljung-Box con retardos de 12, 24, 36 y 48, que se muestra en Anexo N° 3, los residuales son ruido blanco, puesto que los valores de p para los retardos son superiores a 0.05, aceptamos la hipótesis nula.

Prueba de la normalidad de los residuos

H_0 : Los residuos se distribuyen según una distribución normal

H_a : Los residuos no se distribuyen según una distribución normal

Gráfico N° 11: Prueba de normalidad de los residuos para la serie del número de unidades de transfusión de sangre.



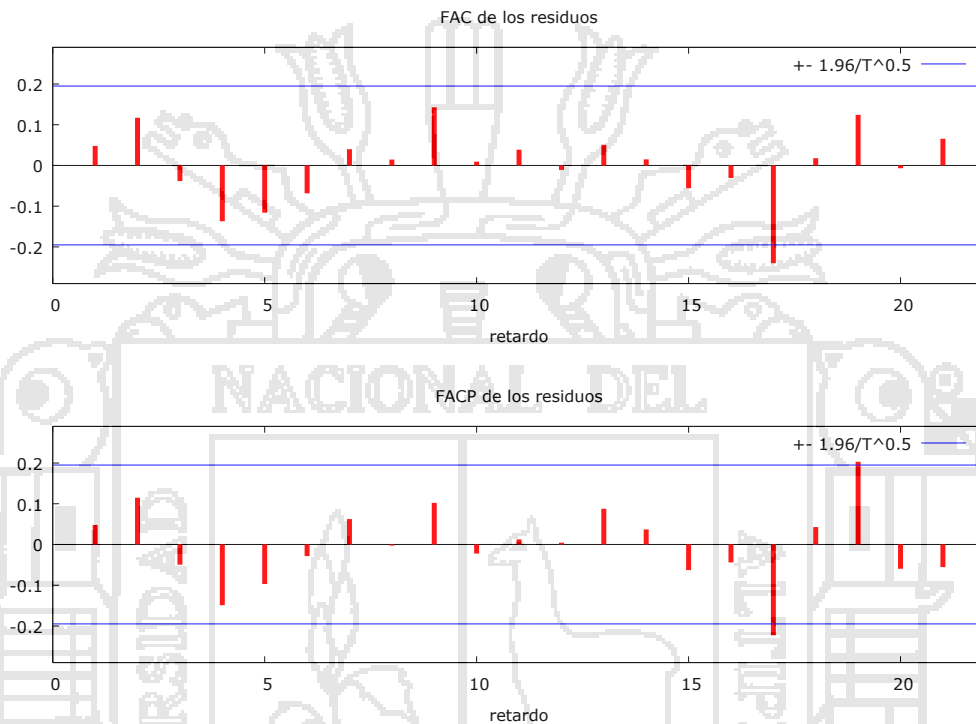
Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

En el gráfico N° 11, Prueba de normalidad de los residuos para la serie del número de transfusiones de sangre ha sido obtenida de software libre Gretl 1.10.1.

Observando el gráfico N°11, el valor de $P > 0.05$, como el valor de P es mayor se acepta la hipótesis nula, confirmamos que los residuales se distribuyen según 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 SARIMA (0, 1,1) (1, 1,0)₁₂ es adecuado para realizar pronósticos.

Gráfico N° 12: FAC Y FACP de los residuos para la serie del número de transfusiones de sangre.



Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

Se observa que en el gráfico N° 12, la función de autocorrelación para los residuales del modelo SARIMA (0, 1,1,) (1, 1,0)_[12], obtenido del software libre Gretl 1.10.1

Para la propuesta de un nivel de significancia $\alpha = 0.05$, por lo menos el 95% de correlación 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.

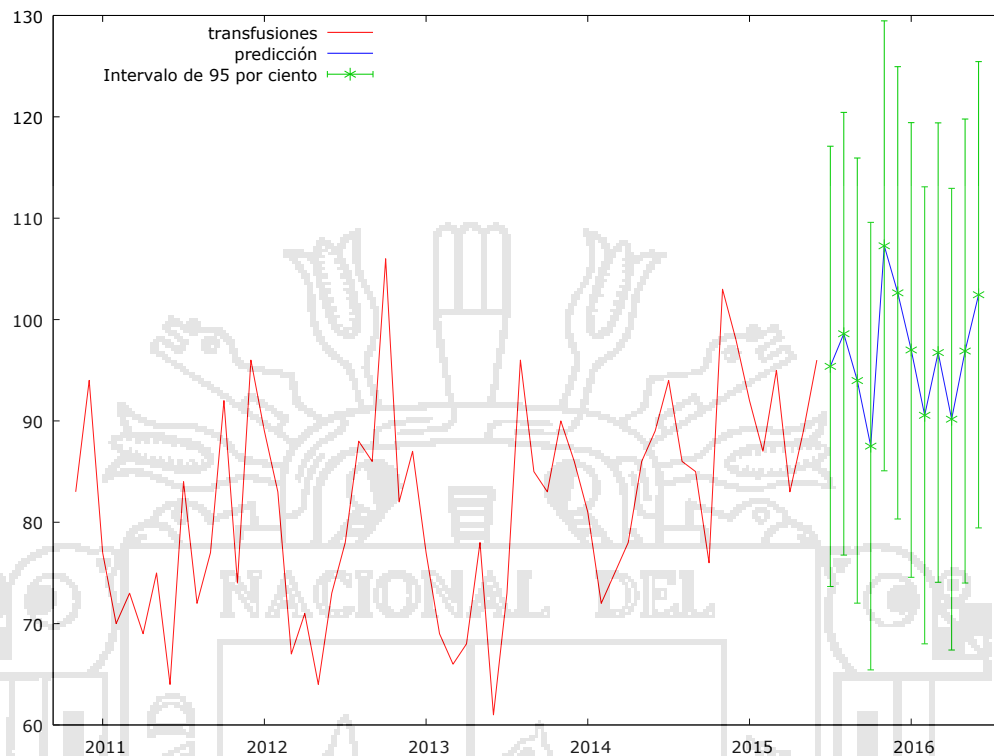
4.1.4. FASE DE REALIZACIÓN DE PRONÓSTICO CON EL MODELO

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 unidades de transfusión de sangre del Hospital Regional Manuel Núñez butrón- Puno.

Los valores previstos para el número de unidades de transfusión de sangre del 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 futuras, muestra una predicción constante ya que el comportamiento es aleatorio dentro de los límites de confianza, Los límites de confianza indican la zona en la que razonablemente estará la futura observación.

Gráfico N° 13: Serie del número de unidades de transfusión de sangre y la serie de pronóstico



Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

Se observa que en el gráfico N° 13, serie del número de unidades de transfusión de sangre y la serie de pronóstico ha sido obtenido del software libre Gretl 1.10.1.

Se observa en el gráfico N° 13, la predicción de número de unidades de transfusiones de Sangre del Hospital Regional Manuel Núñez Butrón-Puno, el número de unidades de transfusión de sangre está proyectada para 12 meses, con el modelo SARIMA (0, 1,1,) (1, 1,0) [12], cuya ecuación de pronóstico es: $\hat{Y}_t = Y_{t-1} + Y_{t-2} - 0.36Y_{t-12} - 0.36Y_{t-13} + \varepsilon_t + 0.89\varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}$ utilizando software libre Gretl 1.10.1.

Tabla N° 7: Pronóstico de la serie del número de unidades de transfusión de sangre.

Pronósticos con el Modelo SARIMA(0,1,1)(1,10) _[12]				
para intervalos de confianza 95%, $z(0.025)=1.96$				
Año	Mes	Pronostico N° de unidades de transfusión de sangre	Intervalos de confianza	
			Inferior	Superior
2015	Julio	96	73.662644	117.100746
	Agosto	99	76.519464	120.430384
	Septiembre	94	72.692653	115.864789
	Octubre	88	65.150172	109.581954
	Noviembre	107	85.628715	129.318599
	Diciembre	103	80.870042	124.816513
2016	Enero	97	75.112113	119.313682
	Febrero	91	68.506815	112.962018
	Marzo	97	74.532616	119.240014
	Abril	90	67.951531	112.909708
	Mayo	97	74.566077	119.773643
	Junio	102	79.963585	125.419173

Fuente: Elaboración propia en base a datos del HRMNB-PUNO (Anexo 01)

En la tabla N°7, Pronostico de la serie del número de unidades de transfusión de sangre, se ha obtenido del software Gretl 1.10.1.

Los resultados proyectados en la tabla N° 7 y el grafico N°11 del número de unidades de transfusión de sangre del Hospital Regional Manuel Núñez Butrón- Puno se muestran pronósticos con un límite inferior y superior del intervalo de confianza del 95%, entonces en el mes de diciembre del año 2015, se tendrá la cantidad de 103 transfusiones de sangre en el Hospital Regional Manuel Núñez Butrón Puno.

CONCLUSIONES

- ❖ 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: **SARIMA** (0, 1,1) (1, 1,0)¹². Cuya ecuación de pronóstico es: $\hat{Y}_t = Y_{t-1} + Y_{t-2} - 0.36Y_{t-12} - 0.36Y_{t-13} + \varepsilon_t + 0.89\varepsilon_{t-1} - \varepsilon_{t-12} + \varepsilon_{t-13}$
- ❖ Los resultados predictivos evaluados para la serie de número de transfusiones de sangre 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 unidades de transfusión de Sangre.
- ❖ Para la predicción de la serie analizada se hizo, uso del mejor modelo estimado y se obtuvieron predicciones para el año 2015-I, comparando con los datos existentes y así comprobar el mínimo error.

RECOMENDACIONES Y SUGERENCIAS

- En primer lugar se recomienda usar de modo general la metodología de Box-Jenkins, que fundamentan la teoría de procesos estocásticos, cuya naturaleza es caracterizada mediante un modelo.
- En el proceso de estimación se recomienda usar las herramientas necesarias para comprobar la estacionariedad e invertibilidad del proceso, de preferencia la estacionariedad.
- Evitar la sobre diferenciación y parametrización, debido a que nos conducen a obtención de modelos erróneos.
- Comparar el grado de eficiencia de la predicción de serie temporal entre redes neuronales y series de tiempo utilizando el software libre Gretl 1.10.1.

BIBLIOGRAFÍA

ANDERSON, O. D. (1985), "*Times Series Análisis and Forecasting*". 1ra Edición Editorial.

AZNAR, J. Z. (2001). "*Análisis de series de tiempo para econometría matemática aplicada*, Editorial Cooperativas, España.

CRUZ M, M. & LLANOS C, M. (2006). "*Modelos Univariantes para describir y pronosticar la afluencia de turistas en la provincial de Puno, Periodo 1995-2005*". Puno.

FERNANDEZ, S. R. (2006). "*Metodología de la Investigación*", Editorial McGRAW-HILL, México.

HANKE, J. E. (2006). "*Pronóstico de Negocios*", Editorial Prentice Hall, México.

OLIVERA, G. M. (1985). "*Time Series Analysis Forecasting and Control*", Editorial Holden-Day, San Francisco.

GRASS, G. (2013). "*Guía práctica para la elaboración de tesis y trabajos de investigación*", Editorial Freire, Riobamba

URIEL, J. E. (1985). "*Análisis de Series de Tiempo Modelos ARIMA*", Editorial Freire, España.

YUPANQUI, C. M. (2010). "*Como Hacer Tesis de investigación*", Editorial Yupanqui, Lima.

REFERENCIAS DE TESIS

CARVAJAL MARTINEZ D. (2011), *“Modelo de series de Tiempo para realizar predicciones de transfusiones sanguíneas en los pacientes ingresados al Hospital Verdi Cevallos en España 2008-2012”*. Tesis de Maestría, Universidad Técnica de Manabí.

CUTIPA PERCCA, Y. (2010). *“Modelo de Series de Tiempo para predecir las atenciones por emergencia, según servicios del Hospital Regional Manuel Núñez Butrón - Puno 2004 - 2007.”* Tesis Pregrado, Universidad Nacional del Altiplano, Puno, Perú.

CENTENO TEVÉS, D. (2013). *“Modelo Univariante para estimar el proceso de atenciones en el Hospital de apoyo Sandia 2013.”* Tesis Pregrado, Universidad Nacional del Altiplano, Puno, Perú.

VALDEZ TICAHUANCA, Y. (2015), *“Modelo de predicción mensual de las Intervenciones quirúrgicas; “aborto” en el Hospital Regional Manuel Núñez-Puno 2006 -2015-I.”* Tesis de Pregrado, Universidad Nacional del Altiplano, Puno, Perú.

CITAS ELECTRÓNICAS

ARELLANO, M. (2008). *Modelos de Series de Tiempo con aplicaciones en la industria aerocomercial*. [En línea]. Consultado: [21, febrero, 2015]
Disponible en:
<http://www.ccee.edu.uy/ensenian/catectr/material/arima_1.pdf>

CARDENAS, ANTÚNEZ, I. (2011). *Técnicas econométricas para el pronóstico de series de tiempo aplicaciones empresariales*. [En línea]. Consultado: [19, marzo, 2015] Disponible en:
<<http://www.ccee.edu.uy/ensenian/catmetec/material/TecnicasEconometricas-SeriesdeTiempo.pdf>>

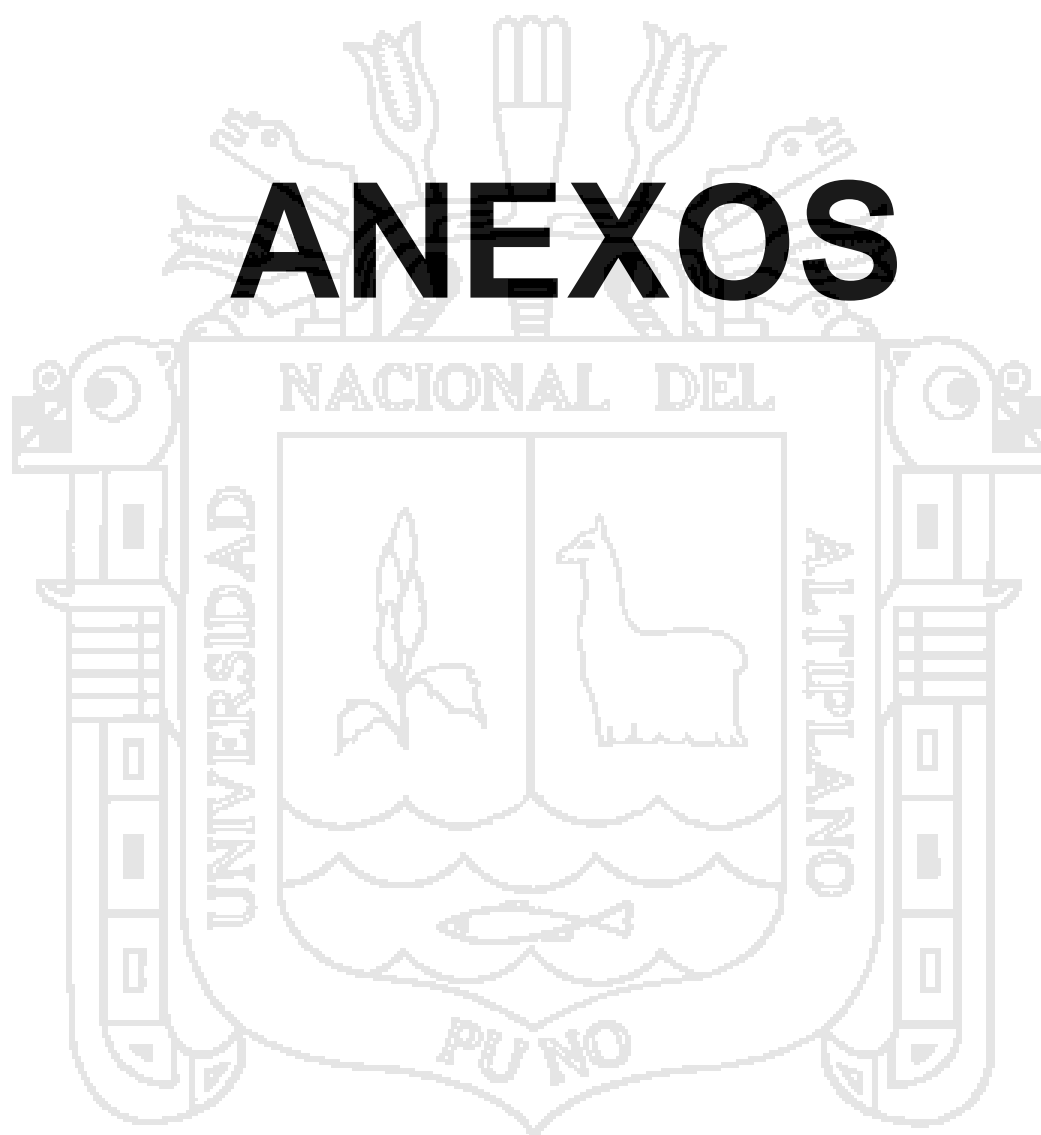
CAPARROS, R. (2010). *Análisis Clásico De Series Temporales*. [En línea]. Consultado: [17, marzo, 2015] Disponible en:
<<http://www.webpersonal.uma.es/antonio/series.pdf>>

DOBERSONW, M. (2012). *Análisis Clásico De Series Temporales*. [En línea]. Consultado: [19, abril, 2015] Disponible en:
<<https://clubmacroeconomiaucv.files.wordpress.com/2012/04/actividad-modulo-2.pdf>>

INEI, (2012). *Desestacionalización de series de tiempo económicas*. [En línea]. Consultado: [12, mayo 2015] Disponible en:
<http://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib0514/Libro.pdf>

OMS, (2015) .Sangre y compontes de grupos Sanguíneos. [En línea]. Consultado: [06, Junio, 2015] Disponible en:
<http://www.who.int/topics/blood_transfusion/es/>.

ANEXOS



ANEXO 1: Serie original del número de unidades de transfusión de sangre en el Hospital Regional Manuel Butrón Puno 2006-2015-I.

MES/AÑO	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
ENERO	66	90	68	62	73	77	89	77	81	92
FEBRERO	62	57	64	67	65	70	83	69	72	87
MARZO	54	66	62	83	84	73	67	66	75	95
ABRIL	83	50	63	69	65	69	71	68	78	83
MAYO	50	68	61	78	76	75	64	78	86	89
JUNIO	61	80	72	87	83	64	73	61	89	96
JULIO	64	66	96	81	89	84	78	73	94	
AGOSTO	56	74	74	86	75	72	88	96	86	
SEPTIEMBRE	64	67	65	90	79	77	86	85	85	
OCTUBRE	57	65	56	82	87	92	106	83	76	
NOVIEMBRE	68	71	66	89	83	74	82	90	103	
DICIEMBRE	73	82	85	87	94	96	87	86	98	

ANEXO 2: Media y Varianza anual del número de unidades de transfusiones de sangre del Hospital Regional Manuel Butrón Puno, 2006-2015-I.

Años	Media	Varianza
2006	61.67	97.333
2007	69.667	118.06
2008	69.333	126.06
2009	80.127	86.065
2010	79.452	82.167
2011	76.833	88.727
2012	81.091	133.53
2013	77.667	113.15
2014	85.25	91.477
2015	90.333	24.667

Anexo 3: Valores de los residuales del Modelo SARIMA (0, 1,1) (1, 1,0)₁₂ para el número de unidades de transfusión de Sangre del Hospital Regional Manuel Núñez Butrón Puno, que se obtuvieron con la ayuda de software libre Grelt 1.10.1.

RETARDO	FAC	FACP	Estad-Q.	[valor p]
1	0.061	0.061		
2	0.1203	0.117		
3	-0.0492	-0.064	2.1662	[0.141]
4	-0.1559	-0.1670 *	4.7724	[0.092]
5	-0.1116	-0.0842	6.121	[0.106]
6	-0.0351	0.0136	6.2561	[0.181]
7	0.0665	0.0827	6.7458	[0.240]
8	0.0333	-0.0035	6.8698	[0.333]
9	0.0784	0.0258	7.5644	[0.373]
10	0.0073	-0.0103	7.5705	[0.477]
11	0.0083	0.0169	7.5785	[0.577]
12	-0.0245	-0.0017	7.6488	[0.663]
13	0.0561	0.0798	8.021	[0.711]
14	0.0424	0.0492	8.2363	[0.766]
15	-0.047	-0.074	8.504	[0.809]
16	-0.0189	-0.0374	8.5478	[0.859]
17	-0.2375 **	-0.2164 **	15.5362	[0.414]
18	0.0021	0.0514	15.5367	[0.486]
19	0.1415	0.2256 **	18.0782	[0.384]
20	0.0271	-0.0297	18.1729	[0.444]

Anexo 4: estimación de los parámetros de los modelos con la ayuda de software libre Gretl 1.10.1

Modelo 1: SARIMA (0, 1,1) (, 1, 0,0), usando las observaciones 2006:02-2015:06 (T = 113)

Variable dependiente: (1-L) transfusion

Desviaciones típicas basadas en el Hessiano

	<i>Coficiente</i>	<i>Desv. Típica</i>	<i>z</i>	<i>Valor p</i>	
Phi_1	0.320745	0.0921739	3.4798	0.0005	***
theta_1	-0.863127	0.0543181	-15.8902	<0.0001	***
Media de la vble. dep.	0.265487		D.T. de la vble. dep.	12.22557	
media innovaciones	1.252276		D.T. innovaciones	9.602706	
Log-verosimilitud	-417.2293		Criterio de Akaike	840.4587	
Criterio de Schwarz	848.6408		Crit. de Hannan-Quinn	843.7789	

Modelo 2: SARIMA (0, 1,1) (1, 1,0), usando las observaciones 2007:02-2015:06 (T = 101)

Variable dependiente: (1-L)(1-Ls) transfusion

Desviaciones típicas basadas en el Hessiano

	<i>Coficiente</i>	<i>Desv. Típica</i>	<i>z</i>	<i>Valor p</i>	
Phi_1	-0.361862	0.10129	-3.5725	0.0004	***
theta_1	-0.894891	0.0598816	-14.9444	<0.0001	***
Media de la vble. dep.	-0.168317		D.T. de la vble. dep.	15.82028	
media innovaciones	-0.119678		D.T. innovaciones	11.08250	
Log-verosimilitud	-387.9949		Criterio de Akaike	781.9898	
Criterio de Schwarz	789.8352		Crit. de Hannan-Quinn	785.1658	