

**UNIVERSIDAD NACIONAL DE SAN ANTONIO ABAD DEL
CUSCO
ESCUELA DE POSGRADO
MAESTRÍA EN ESTADÍSTICA**



T E S I S

**MODELO DE REDES NEURONALES ARTIFICIALES PARA EL
PRONÓSTICO DEL NÚMERO DE VISITANTES EXTRANJEROS A
MACHU PICCHU EN COMPARACIÓN CON LA METODOLOGÍA DE
BOX Y JENKINS**

Presentado por:

Br. Henry Lazo Chuquiwayta

Para Optar al Grado Académico de:

Maestro en Estadística

Asesora:

Mgt. Yetsabel Aucaille Quispe

CUSCO – PERÚ

2019

INDICE GENERAL

AGRADECIMIENTO	i
DEDICATORIA	ii
PRESENTACIÓN	iii
RESUMEN	iv
ABSTRACT	v
CAPÍTULO I	1
PLANTEAMIENTO DEL PROBLEMA	1
1.1. Situación Problemática	1
1.2. Formulación del Problema	2
1.3. Justificación de la Investigación	3
1.4. Objetivos de la Investigación	3
1.4.1. Objetivo General.....	3
1.4.2. Objetivos Específicos.	3
CAPÍTULO II	4
MARCO TEÓRICO CONCEPTUAL	4
2.2. Antecedentes de la Investigación	4
2.2.1. Antecedente Internacional.	4
2.2.2. Antecedente Nacional.....	6
2.2.3. Antecedente Regional.....	7
2.2. Base Teórica	8
2.2.1. Series de tiempo.	8
2.2.2. Redes Neuronales.	43
2.2.2.1. Ecuaciones matemáticas con el modelo ARIMA y SARIMA	44
2.2.2.2. Criterio de información Akaike (AIC)	45
2.2.3. Neuronas Biológicas.....	45
2.3. Marco Conceptual	68

CAPÍTULO III	71
HIPÓTESIS Y VARIABLES	71
3.1. Hipótesis	71
3.1.1.Hipótesis general	71
3.1.2.Hipótesis específicas.....	71
3.2. Operacionalización de variables	72
CAPÍTULO IV.....	73
METODOLOGÍA.....	73
4.1. Tipo y Diseño de Investigación	73
4.2. Población	73
4.3. Muestra no probabilística	73
4.4. Técnicas de recolección de datos e información	73
4.5. Métodos de Tratamiento de Datos.....	74
CAPÍTULO V	79
RESULTADOS.....	79
5.1. Procesamiento, análisis e interpretación de los datos	79
5.1.1. Método de Box y Jenkins.	79
5.1.1.1. Identificación del modelo.	79
5.1.1.2. Estimación del modelo.	86
5.1.1.3. Evaluación del modelo	87
5.1.1.4. Realización de pronósticos con el modelo.	88
5.1.2. Método de redes neuronales.	89
5.1.2.1. Escalamiento de los datos.....	89
5.1.2.2. Patrones de entrenamiento y prueba.....	90
5.1.2.3. Topología de la red neuronal artificial.	90
5.1.2.4. Determinación de las entradas.....	90
5.1.2.5. Determinación del número de neuronas en la capa oculta.	91
5.1.2.6. Algoritmo de entrenamiento.....	91
5.1.2.7. Selección de los pesos iniciales.	91
5.1.2.8. Ajuste o entrenamiento de la red neuronal artificial.....	91

5.1.2.9. Pronostico.....	91
5.1.3. Modelos de redes neuronales para pronóstico.....	91
CONCLUSIONES	95
RECOMENDACIONES	96
BIBLIOGRAFÍA	97
- ANEXOS -.....	101
- ANEXO 1 -.....	102
- ANEXO 2 -.....	105
- ANEXO 3 -.....	107

ÍNDICE DE TABLAS

Tabla 1:Patrones teóricos de FAC y FACP	37
Tabla 2:Operacionalización de variables.....	72
Tabla 3:Comparación de modelos tentativos para la serie número mensual de visitantes extranjeros.....	82
Tabla 4:Coeficientes estimados para el modelo seleccionado	86
Tabla 5:Valores reales y pronosticados del número mensual de visitantes extranjeros.....	88
Tabla 7:Modelo elegido de la Metodología Box-Yenkins (ARIMA) y Redes Neuronales Artificiales (RNA)	93

ÍNDICE DE FIGURAS

Figura 1: Esquemático de Neuronas Biológicas. 46	
Figura 2: Modelo de una Neurona.....	48
Figura 3: Función Paso	49
Figura 4: Función Lineal.....	50
Figura 5: Función Rampa.....	50
Figura 6: Función Logística.....	51
Figura 7: Función Tangente Hiperbólica.....	52
Figura 8: Función Gaussiana.....	53
Figura 9: Ejemplo de una arquitectura de red.....	56
Figura 10: Red Neuronal de alimentación adelante con una capa oculta para la predicción de series de tiempo.....	62
Figura 11: Conexión entre tres neuronas.....	65
Figura 12: Backpropagation.....	66
Figura 13: Gráfico de secuencia de la serie del número mensual de visitantes extranjeros.....	79
Figura 14: Diagrama de caja de la serie número mensual de Visitantes extranjeros.....	80
Figura 15: Diagrama de cajas por mes de los años de la serie número visitantes de extranjeros.....	81
Figura 16: Función de Autocorrelación de la serie número mensual de visitantes extranjeros.....	82
Figura 17: Diagnóstico de los residuos del modelo ajustado.....	87
Figura 18: Valores ajustados y pronosticados con límites de confianza al 95% el modelo ARIMA (1,0,1)(0,1,1) ₁₂ para serie número de visitantes extranjeros.....	88
Figura 19: Valores ajustados y pronosticados por el modelo de red neuronal artificial para serie número de visitantes extranjeros.....	93
Figura 20: Serie de tiempo real, serie con modelo arima, serie con modelo de red neuronal y pronósticos del modelo arima-red neuronal artificial.....	91

AGRADECIMIENTO

Este trabajo lo dedico en especial a mi Madre Leandra, que me supo apoyar en todo momento y dar todo el ánimo para seguir hasta concluir toda la Maestría.

Así mismo a mis hermanos (Elisabet y Carlos) y a todos mis sobrinos para que lo tomen como ejemplo.

Gracias a todos los docentes de la Maestría en Estadística por sus enseñanzas académicas en toda mi formación; razón que decían uno deja de estudiar cuando ya no está presente.

DEDICATORIA

*A Dios todo poderoso que me guía en
toda mi vida, dándome salud, sabiduría
y bendiciendo en mi hogar.*

*En memoria de mi tía Adela Márquez
Palomino, mi tío Max G. Loayza Mariaca que
están al lado del Señor.*

“A mi docente **Gabrielle Frish** que está al lado del Señor”.

*A todos que los conocí y que ya no se
encuentran presentes, sé que todos
llegaremos a esa última etapa de
nuestras vidas.*

PRESENTACIÓN

SEÑOR DIRECTOR DE LA ESCUELA DE POSGRADO.

SEÑOR DIRECTOR DE LA MAESTRIA EN ESTADISTICA.

SEÑORES MIEMBROS DEL JURADO:

De acuerdo al Reglamento establecido por la Escuela de Posgrado de la Universidad Nacional de San Antonio Abad del Cusco, Maestría en Estadística ha sido elaborada la Tesis Titulado:

“MODELO DE REDES NEURONALES ARTIFICIALES PARA EL PRONÓSTICO DEL NÚMERO DE VISITANTES EXTRANJEROS A MACHU PICCHU 2002-2018 EN COMPARACIÓN CON LA METODOLOGÍA DE BOX Y JENKINS”, la misma

que pongo a vuestra consideración para la revisión y evaluación correspondiente. Este trabajo de investigación tiene como finalidad usar el modelo adecuado para la precisión de los datos y el pronóstico de los mismos la misma que será de utilidad para las entidades dedicadas al sector turismo en la región.

El presente trabajo de investigación está estructurado en capítulos establecidos en la siguiente forma:

El primer capítulo se describe el planteamiento del problema, situación problemática, formulación del problema, justificación de la investigación y objetivos.

El segundo capítulo se ha desarrollado el Marco Teórico Conceptual, antecedentes de la investigación, base teórica, redes neuronales y neuronas biológicas.

En el tercer capítulo esta descrita la hipótesis y variables: operacionalización de variables.

En el cuarto capítulo se desarrolla la metodología empleada para el desarrollo de la tesis.

El quinto capítulo se presenta los resultados obtenidos de la tesis.

Para finalizar mi investigación se tiene las conclusiones, recomendaciones y la bibliografía utilizada.

Henry Lazo Chuquiwayta

RESUMEN

El objetivo fundamental fue comparar la metodología de Box-Jenkins, que se utiliza para pronosticar con series de tiempo, y el método de redes neuronales artificiales, que es un método de la inteligencia artificial. A la serie número mensual de visitantes extranjeros al santuario histórico de Machu Picchu de Cusco 2002-2018 se le aplicó la metodología de Box-Jenkins y el método de redes neuronales artificiales. Particularmente, se utiliza el modelo de red neuronal tipo Perceptrón Multicapa que tuvo seis entradas o retrasos de la serie (Y_{t-1} , Y_{t-2} , Y_{t-3} , Y_{t-4} , Y_{t-12} , Y_{t-24}), donde se consideró cuatro neuronas (nodos) en la capa oculta y una neurona (nodo) en la capa de salida, el cual tuvo una capacidad de predicción más precisa con un MAPE de 4.41% a comparación del modelo ARIMA que tiene un MAPE de 7.66%. Los pronósticos para el primer semestre del 2018 realizados por el modelo de red neuronal se parecen más al patrón histórico de la serie. Los datos fueron analizados utilizando el software R-project.

Palabras claves: Pronóstico, red neuronal artificial, Perceptrón Multicapa, metodología Box-Jenkins, ARIMA.

ABSTRACT

The main objective was to compare the Box-Jenkins methodology, which is used to forecast with time series, and the method of artificial neural networks, which is a method of artificial intelligence. The monthly the series of foreign visitors to the historic sanctuary of Machu Picchu of Cusco 2002-2018 was applied the Box-Jenkins methodology and the method of artificial neural network. Particularly, the Multilayer Perceptron type neural network model that had six inputs or delays of the series $(Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}, Y_{t-12}, Y_{t-24})$, where we considered four neurons (nodes) in the hidden layer and a neuron (node) in the output layer, which had a more accurate prediction capability with a MAPE of 4.41% compared to the ARIMA model that has a MAPE of 7.66%. The forecasts for the first half of 2018 made by the neural network model are more similar to the historical pattern of the series. The data was analyzed using the software R-project.

Keywords: Prognosis, artificial neural network, Multilayer Perceptron, Box-Jenkins methodology, ARIMA.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1. Situación Problemática

El modelamiento de series temporales, según Box y Jenkins & Reinsel (1994), “es un área de interés para muchas disciplinas y en el que muchos esfuerzos se han dedicado para el desarrollo de nuevos métodos”.

El objetivo es proveer una función matemática de una serie temporal, que permita capturar total o parcialmente, las características más relevantes del fenómeno real, a partir de la información contenida en los datos. En la literatura se han propuesto diversos modelos orientados a la representación de series temporales (Hanke, J.E. & Wichern, D.W. 2006). La mayoría de los casos reales presentan irregularidades en las cuales no se pueden utilizar los métodos tradicionales de la estadística, lo cual nos conduce a dar solución a este problema mediante nuevos procedimientos. Particularmente, se presentará el método de redes neuronales artificiales, que han demostrado su capacidad para resolver problemas complejos en diversas áreas de series de tiempo indican Hyndman, J.R, & Khandakar, Y.(2008). Esta metodología será aplicada a la serie de tiempo del número total mensual de visitantes extranjeros al Santuario Histórico de Machu Picchu del Cusco y se procederá a obtener sus pronósticos mediante el método de Redes Neuronales Artificiales. Asimismo, se aplica las Metodologías de Box y Jenkins, para realizar las estimaciones y pronósticos. Finalmente, los resultados de ambos modelos serán comparados.

1.2. Formulación del Problema

La predicción y el modelado de series temporales es de gran interés práctico, pues permite conocer valores futuros con un margen de error de una serie basándose en sus valores pasados, lo cual podría facilitar la toma de decisiones y la planificación en muchas áreas de la ciencia.

La metodología de Box y Jenkins son utilizados con mucha frecuencia para predecir series de tiempo. Además, del criterio del investigador, los pasos a realizar dependen del tipo de datos (históricos), que deben ser estacionarios, se analizan los gráficos de correlaciones, se evalúa la adecuación del modelo, se mide el error de predicción, etc. Box, Jenkins, & Reinsel, (1994). En general, estos procedimientos resultan complejos. Por otra parte, la Inteligencia Artificial es relativamente una nueva metodología para predecir y está formado por un conjunto de métodos que intentan imitar en forma artificial las habilidades relacionadas con la inteligencia humana. Uno de estos métodos que trata de simular el pensamiento humano mediante conexión de neuronas es la denominada “Redes Neuronales Artificiales”, que son utilizadas con un relativo éxito para la predicción de series de tiempo. La descripción de ambas metodologías conducen de manera natural a la siguiente pregunta:

¿Es posible que el modelo de redes neuronales artificiales obtenga mejores pronósticos del número de visitantes extranjeros a Machu Picchu 2002-2018 en comparación con la metodología de Box y Jenkins?

1.3. Justificación de la Investigación

En la presente investigación se muestra el uso de modelos de redes neuronales artificiales, aplicadas a las ciencias sociales, en específico al sector del turismo, para probar la utilidad que prestan como modelos de predicción del número de visitantes extranjeros al Santuario Histórico de Machu Picchu en base a la identificación de patrones de comportamientos pasados (datos históricos). Proveen información sobre el comportamiento futuro del número de visitantes extranjeros al Santuario Histórico de Machu Picchu a las entidades interesadas en el sector turismo, puesto que a partir de pronósticos confiables las entidades podrán tomar mejores decisiones.

1.4. Objetivos de la Investigación

1.4.1. Objetivo General.

Comparar el modelo de redes neuronales artificiales para el pronóstico de número de visitantes extranjeros a Machu Picchu 2002-2018 con la metodología de Box y Jenkins.

1.4.2. Objetivos Específicos.

- a. Determinar el modelo adecuado de Redes Neuronales Artificiales para el pronóstico del número de visitantes extranjeros a Machu Picchu 2002-2018.
- b. Pronosticar el número de visitantes extranjeros a Machu Picchu 2002-2018 mediante el modelo de Redes Neuronales Artificiales y la metodología de Box y Jenkins.
- c. Analizar los resultados obtenidos mediante un modelo de Redes Neuronales Artificiales y la metodología de Box y Jenkins.

CAPITULO II

MARCO TEÓRICO CONCEPTUAL

2.2. Antecedentes de la Investigación

2.2.1. Antecedente Internacional.

En la tesis, Joekes Silvia, Barbosa, Emanuel P., & Robledo, Walter. (2005) sustentan el “modelado y pronóstico de una serie de tiempo contaminada empleando redes neuronales y procedimientos estadísticos tradicionales”, concluye de que al comparar los resultados de la red robustamente entrenada con los obtenidos ajustando el modelo AR(2) se pudieron observar muy pocas diferencias en el ajuste y pronóstico de ambos procedimientos. El modelo estadístico mostró un menor error residual y menor porcentaje de error absoluto en el ajuste, pero la red robustamente entrenada superó al modelo estadístico en los pronósticos. Previo a efectuar la corrección para los valores extremos, el modelo de red logró un buen ajuste de los datos de la serie y un mejor pronóstico que el modelo estadístico corregido. Esto se debe a que el efecto de los valores extremos en las redes neuronales es sólo local en contraposición con lo que sucede con los modelos estadísticos”.

En la tesis Morales Castro, Arturo.& Cruz Torres, Iván. (2008) titulado: “Pronósticos mediante redes neuronales artificiales y modelos ARIMA: el caso de los cetes en México”, concluye según los resultados y las pruebas de hipótesis, tanto de la bondad de ajuste representada por r^2 como las pruebas estadísticas de Akaike y Schwarz, indican que el modelo basado en Redes Neuronales Artificiales es mejor que el modelo ARIMA calculado y usado como

referencia dentro de la muestra y fuera de la muestra, también se logró una ligera mejora expresada por los estadísticos, raíz cuadrada de error medio (RMSE, por sus siglas en inglés) y U de Theil; sin embargo, el error medio absoluto porcentual (MAPE, por sus siglas en inglés) resulto mayor al planteado en la hipótesis, por lo que se considera que es mejor modelo para pronósticos aunque no de manera absoluta. referente al proceso de encontrar un modelo basado en Redes Neuronales sostiene que es más complejo y delicado que el desarrollo de un modelo ARIMA, ya que implica el cálculo de más parámetros, y la estabilización de la red requiere mayores recursos de cómputo y tiempo, pues el cálculo de sus parámetros es, de forma inherente, a prueba y error.

En su artículo, Camilo Santana, Juan. (2006) analiza la “Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana, concluye que las predicciones obtenidas de las variaciones en el índice de precios al consumidor (IPC) a través de redes neuronales tienden a ser más precisas que las originales por metodologías tradicionales. Adicionalmente, la utilidad de las redes neuronales, en combinación con otras redes o metodologías tradicionales, se mostró efectiva en el proceso de predicción en términos del cuadrado medio del error (MSE). Finalmente, las redes neuronales sin la neurona de tendencia, en la mayoría de casos exhibieron las menores medidas de error en la predicción, en comparación con la metodología utilizada que propone la tendencia como una neurona de entrada”.

En su tesis, Gómez López , Williams. (2010) presenta el “Pronóstico de una serie temporal de demanda máxima de energía eléctrica usando redes neuronales" donde llega a la conclusión de que las predicciones que proporcionó

la red neuronal, en los cinco primeros meses, tuvieron un error menor que los valores proporcionados por el modelo de serie de tiempo. Cabe destacar que conforme se va aumentando el periodo al cual se desea pronosticar, el error absoluto de la predicción usando series de tiempo incrementa, además de que existirá un periodo donde a partir de este, los valores que la serie proporcione serán constantes. En la red neuronal, no se conoce con exactitud el comportamiento que pueda tener la red conforme se aumente el número de periodos que se desea pronosticar. Al desarrollar el modelo de redes neuronales no existe previamente una estructura preliminar para poder resolver el problema que se plantea; pues bien solo se afirma de la existencia de una estructura (al menos) el cual puede dar respuestas adecuadas al problema. Las redes neuronales proporcionan buenas predicciones comparados con las predicciones del modelo de serie de tiempo.

2.2.2. Antecedente Nacional.

En la tesis, Cárdenas Garro, José. Antonio. (2015) presenta el “Pronósticos y comparación de una serie de tiempo con cambios estructurales mediante la red neuronal artificial de Retropropagación Resiliente y modelos no lineales” concluye que el mejor modelo que nos otorga los mejores pronósticos para la serie de tiempo: Número de peruanos retornantes según año de regreso mediante medio de transporte aéreo, es el segundo modelo no paramétrico, donde se utilizó la metodología de Redes Neuronales Artificiales por el método de Retropropagación Resiliente para los pronósticos del año 2012, por lo que concluye que la serie de tiempo con cambios estructurales la mejor metodología para el caso de pronósticos es la de Redes Neuronales Artificiales”.

En la tesis, Esteba M, N. (2011) investiga las “técnicas de redes neuronales artificiales para predecir el ingreso de extranjeros al Perú por el Aeropuerto Internacional Jorge Chávez”, llegó a la conclusión de que el potencial de las redes neuronales artificiales como herramienta para el cálculo de previsiones de ingreso de extranjeros. Gracias a su capacidad de aprender a partir de los datos que se le proporcionó, sin hacer suposiciones sobre los modelos y relaciones que subyacen en la serie, pueden ser aplicadas directamente, lo que facilita considerablemente el proceso de cálculo de previsiones”.

2.2.3. Antecedente Regional

La tesis, Mamani Barrios, Eulios. (2014) presenta, “Redes neuronales artificiales para el pronóstico del número de visitantes extranjeros a la reserva nacional del Titicaca”, concluye la arquitectura de tipo Perceptrón Multicapa para el pronóstico, tuvo como entradas siete valores retrasados de la serie (y_{t-1} , y_{t-2} , y_{t-3} , y_{t-4} , y_{t-12} , y_{t-13} , y_{t-24}), los cuales fueron seleccionados por ensayo y error en donde se contó con el apoyo de los correlogramas simple y parcial para el óptimo número de entradas, cuatro neuronas (nodos) en la capa oculta que fueron también elegidos por ensayo y error, y una neurona en la capa de salida (y_t). En el proceso de entrenamiento se estimó 37 pesos (parámetros) del modelo seleccionado con un MAPE de 19.13% en el ajuste, los resultados mostraron que el modelo seleccionado tuvo una capacidad de predicción precisa en la muestra de validación con un MAPE de 7.23% cifra menor a comparación de los modelos restantes considerados. Los pronósticos de 12 meses en adelante realizados presentan un comportamiento estable.

Otra tesis relacionada a Redes Neuronales es presentado por Valencia Toledo, Alfredo & Quispe Huilcas, Raul (2009), que titula “Modelos no Lineales de Redes Neuronales Artificiales Basadas en la Predicción de Series de Tiempo”, y aplicaron para predecir series mensuales en el área de economía. Llegando a las conclusiones: los modelos de redes neuronales artificiales presentan mejores resultados en el ajuste de predicción, al utilizar una muestra pequeña; nos muestra la eficacia en el ajuste de predicción que tiene los modelos de Redes Neuronales Artificiales en las series de tiempo frente a los modelos lineales; una ventaja que brinda esta metodología es que a diferencia de los métodos econométricos clásicos, el modelo de las Redes Neuronales Artificiales nos permite diseñar un modelo que se ajuste a los datos; la utilización de las Redes Neuronales Artificiales no necesariamente desplazara a los métodos econométricos clásicos, por lo contrario coadyuvara como un nuevo método de predicción.

2.2. Base Teórica

2.2.1. Series de tiempo.

“Es un conjunto de observaciones ordenadas según una característica cuantitativa de un fenómeno individual en diferentes momentos del tiempo, en el cual las observaciones son realizadas” (Anderson O, D.1985).

“Una serie de tiempo consta de datos que se reúnen, registran u observan sobre incrementos sucesivos de tiempo” (Hanke J, E.1996).

“Una serie de tiempo es una secuencia cronológica de observaciones de una variable en particular” (Bowerman B, L. & O’connell R, T.1993).

Un modelo clásico de series de tiempo, supone que la serie Y_1, \dots, Y_T puede ser expresada en sus componentes:

Modelo aditivo: $Y(t) = T(t) + E(t) + C(t) + A(t)$

Modelo multiplicativo: $Y(t) = T(t) \cdot E(t) \cdot C(t) \cdot A(t)$

Componentes de una serie temporal

“En el análisis de series de tiempo de datos, una tentación inmediata consiste en intentar explicar o contabilizar el comportamiento de las series. La descomposición clásica es un método que se basa en la suposición de que se pueden descomponer en componentes como tendencia, ciclo, estacionalidad e irregularidad. Una predicción se hace mediante la combinación de las proyecciones de cada componente individual”. (Hanke J,E. 1996)

a. Tendencia

La tendencia es un movimiento de larga duración que muestra la evolución general de la serie en el tiempo. Es un movimiento que puede ser estacionario o ascendente o descendente, y su recorrido, una línea recta o una curva.

b. Componente cíclico o variación cíclica

El componente cíclico es un conjunto de fluctuaciones en forma de onda o ciclos, de más de un año de duración. El ciclo sugiere la idea de que este

tipo de movimiento se repite cada cierto periodo con características parecidas.

c. **Componente estacional o variación estacional**

Se habla de este tipo de variaciones usualmente cuando el comportamiento de la variable en el tiempo en un periodo está relacionado con la época o un periodo particular, por lo general en el espacio cronológico presente.

El componente estacional es un patrón de cambio que se repite a sí mismo año tras año. Se encuentran típicamente en los datos clasificados por trimestre, mes o semana.

d. **Componente aleatorio o variación residual**

El componente aleatorio mide la variabilidad de las series de tiempo después de que se retiran los otros componentes. Contabiliza la variabilidad aleatoria en una serie de tiempo ocasionada por factores imprevistos y no recurrentes.

La mayoría de los componentes irregulares se conforman de variabilidad aleatoria. Sin embargo, ciertos sucesos a veces impredecibles como huelgas, cambios de clima (inundaciones, sequías o terremotos), elecciones, conflictos armados o la aprobación de asuntos legislativos, pueden causar irregularidades en una variable. (Hanke J, E. 1996)

Análisis de series de tiempo

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

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

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

- a. **Descripción:** Al tener una serie de tiempo, el primer paso en el análisis es graficar los datos y obtener medidas descriptivas simples de las propiedades principales de la serie.
- b. **Explicación:** Cuando las observaciones son tomadas sobre dos o más variables, es posible usar la variación en una serie para explicar la variación en las otras series.
- c. **Predicción:** Dada una serie de tiempo se intenta predecir los valores futuros de la serie. Este es el objetivo más frecuente en el análisis de series de tiempo.
- d. **Control:** Si una serie de tiempo se genera por mediciones de calidad de un proceso, el objetivo del análisis puede ser el control del proceso.

Estimación de la tendencia

La tendencia se estima de la siguiente manera:

✓ **Método de los mínimos cuadrados ordinarios**

El método de los mínimos cuadrados, se utiliza para calcular la ecuación de una recta o curva de tendencia apropiada. Permite identificar el grado de correlación entre una variable dependiente y una o más variables independientes. El criterio de este método es usar la recta ($Y = a + bX$) cuya suma de los cuadrados de los errores sea mínima.

Estimación de los parámetros:

$$b = \frac{n \sum XY - \sum X \sum Y}{n \sum X^2 - (\sum X)^2}$$

$$a = \frac{\sum X^2 \sum Y - \sum X \sum XY}{n \sum X^2 - (\sum X)^2} \qquad a = \bar{Y} - b\bar{X}$$

Donde:

a : Intersección en el eje vertical.

b : Pendiente de la línea de regresión

n : Número de observaciones.

X : Valores de Y que caen en la línea de tendencia.

Y : Valores de X que caen en la línea de tendencia.

Pronósticos

Los pronósticos son predicciones de lo que puede suceder o esperar, son premisas o suposiciones básicas en que se basa la planeación y la toma de decisiones. El propósito del pronóstico consiste en reducir el margen de incertidumbre, haciendo

el mejor uso de la información que se tiene para guiar las actividades de la empresa hacia el cumplimiento de sus metas y objetivos. De esta forma los pronósticos son particularmente importantes en la asignación del uso de los recursos de la empresa.

Los pronósticos utilizan datos anteriores de una variable para predecir su desempeño futuro. Así mismo, los datos anteriores se dan generalmente en la forma de series de tiempo. Una hipótesis básica en la aplicación de las técnicas de pronóstico es que el desempeño de los datos anteriores continuará ocurriendo en el futuro inmediato.

La necesidad de pronosticar

“Debido a que siempre ha sido cambiante el mundo en el que operan las organizaciones, es frecuente la necesidad de hacer pronósticos. Sin embargo, en los últimos años se incrementó la confianza en las técnicas que abarcan una compleja manipulación de datos” (Hanke J, E.1996).

El avance tecnológico, junto con las técnicas cuantitativas que hacen posible, se han vuelto más que recomendables en las organizaciones modernas; siendo esenciales.

¿Quién requiere hacer pronósticos? Casi cualquier organización, grande o pequeña, pública o privada, utiliza el pronóstico, debido a que casi todas las organizaciones deben planear cómo enfrentar las condiciones futuras de las cuales tiene un conocimiento imperfecto. Además, la necesidad de hacer pronósticos cruza todas las líneas funcionales, lo mismo que todo tipo de organizaciones.

Técnicas de pronósticos

“Se pueden emplear dos técnicas básicas de pronósticos: Las técnicas de pronóstico cualitativas y las técnicas de pronóstico cuantitativas” (Hanke J, E.1996).

✓ **Técnicas de pronóstico cuantitativas**

Se utilizan cuando existen suficientes datos históricos disponibles y cuando se juzga que estos datos son representativos de un futuro desconocido.

Trabajan con modelos cuantitativos o modelos matemáticos que se basan en datos históricos, bajo el supuesto de que son relevantes para el futuro.

Estos modelos se pueden utilizar con series de tiempo.

Pronostico según plazos

Los pronósticos a largo plazo son necesarios para establecer el curso general de la organización para un largo periodo, sirven para tomar decisiones estratégicas y por lo general abarcan de tres a cinco años. Los pronósticos a mediano plazo abarcan de uno a dos años. Los pronósticos a corto plazo se utilizan para diseñar estrategias inmediatas que ayuden en la toma de decisiones, sólo abarcan meses.

“Las técnicas más complejas de Box-Jenkins resultan apropiadas para pronósticos de corto y mediano plazos” (Hanke J,E.1996).

Proceso estocástico

“Sea τ un conjunto arbitrario. Un proceso estocástico es una familia $\{X(t), t \in \tau\}$

tal que para cada $t \in \tau$, $X(t)$ es una variable aleatoria” (Morettin, P. 2002).

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

$$Y_1, Y_2, \dots, Y_t, \dots = \{Y_t / t \in \Gamma, \forall t \in \Gamma\}$$

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

Ruido blanco

Se llama ruido blanco a una sucesión de variables aleatorias (proceso estocástico) con esperanza (media) cero, varianza constante e independientes para distintos valores de t (covarianza nula). Es llamado proceso de ruido blanco siendo un proceso estacionario ε_t .

Una variable ε_t (ruido blanco), cumple las siguientes propiedades:

$$i) E[\varepsilon_t] = 0 \quad \forall t$$

$$ii) E[\varepsilon_t]^2 = \sigma^2 \quad \forall t$$

$$iii) \text{COV}[\varepsilon_t, \varepsilon_s] = E[\varepsilon_t, \varepsilon_s] = 0 \quad \forall t \neq s$$

Estacionalidad

Es una variación repetitiva a lo largo de un intervalo de tiempo.

Se dice que una serie de tiempo es estacional cuando los datos tienen oscilaciones estrictamente periódicas, donde el periodo es igual o inferior al año, por ejemplo la repetición del patrón puede ser cada 3 meses, 6 meses, cada año (12 meses), cada 4 años.

Coeficiente de correlación

Mide el grado de independencia de una variable relacionada con otra variable. Es una cantidad que está entre -1 y +1 ($-1 \leq r \leq 1$), presenta el grado de correlación entre dichas variables; mientras este valor se aproxima a los límites, diremos que la correlación es fuerte.

Se expresa:

$$r = \frac{\text{COV}(x, y)}{\sigma_x \sigma_y}$$

r^2 es el coeficiente de determinación, y es una medida de ajuste de la Regresión. Es determinado por la proporción de la suma de cuadrados del total. Se expresa como:

$$r^2 = \frac{SCR}{SCT} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

El error siempre existirá; en estadística es posible lograr una menor varianza de error σ_e^2 , como lograr un r^2 cercano a los límites.

Varianza y error estándar de ruido blanco

- ✓ La varianza de ruido blanco de la regresión es una medida que sirve para analizar la capacidad explicativa del modelo.

Se expresa como:
$$\sigma_e^2 = \frac{\sum_{p=1}^T (Y_t - \hat{Y})^2}{T - k}$$

Donde:

T: es el número de residuos o errores.

K: es el número de estimadores.

✓ El error estándar o desviación estándar de la regresión está dado por:

$$\sigma_e = \sqrt{\sigma_e^2}$$

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, en forma matemática” (Kendall M, G. & Buckland W, R. 1985).

Es un esquema teórico, generalmente en forma matemática, de un sistema o de una realidad compleja, como la evolución económica de un país, que se elabora para facilitar su comprensión y el estudio de su comportamiento.

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 llamada serie cronológica. A las observaciones de una variable temporal se les denomina “realizaciones”.

Modelo univariante

“Los modelos univariantes en una serie de tiempo $\{Y_t\}$, son todos aquellos que solamente tienen una variable observada en el tiempo. Estos tipos de modelos se expresan en forma polinomial” (Anderson O, D.1985).

Son técnicas univariantes: 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 Box-Jenkins, o modelos ARIMA, pues las técnicas Box-Jenkins constituyen un conjunto más amplio, dentro del cual los modelos ARIMA univariantes son sólo una parte.

a. Modelo univariante no integrado

Los procesos Autorregresivos AR (p), 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.

b. **Modelo univariante integrado**

Son aquellos modelos que se pueden obtener mediante suma o integración de un proceso estacionario. A estos modelos se les denomina también modelos no estacionarios homogéneos.

Los procesos Mixtos Integrados ARIMA(p,d,q), los procesos Estacionales Mixtos Integrados SARIMA(p,d,q)*(P,D,Q), procesos de Medias Móviles Integrado IMA, proceso de Medias Móviles Exponenciales EWMA, y los procesos de Autoagregación, son considerados como los modelos Integrados debido a que si interviene el Grado de Diferenciación y la Estacionalidad de la serie.

Operador de retardo y diferenciación de una serie

Introduciremos a continuación el operador polinomial de retardos L. El operador L determina que:

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

Es decir, el resultado de aplicar el operador L corresponde a la observación en el período anterior de la variable (serie).

Aplicada dos veces sobre la variable Y_t es:

$$L(LY_t) = L^2Y_t = Y_{t-2}$$

En general: $L^kY_t = Y_{t-k}$

La diferencia de una serie es:

$$\Delta Y_t = Y_t - Y_{t-1} = (1 - L)Y_t$$

En general:

$$\Delta^k Y_k = \Delta(\Delta^{k-1} Y_t) = (1-L)^k Y_t$$

Elaboración de modelos: AR, MA, ARMA y ARIMA

Los modelos ARIMA o modelos de promedio móvil autorregresivo integrado son un tipo general del método de Box-Jenkins para series de tiempo estacionarias. Una serie histórica estacionaria es aquella cuyo valor promedio no cambia a través del tiempo. Este grupo incluye a los modelos AR sólo con términos autorregresivos, los modelos MA sólo con términos de promedio móvil y los modelos ARIMA que comprenden tanto términos autorregresivos como de promedio móvil. El método de Box-Jenkins permite al analista seleccionar el modelo que mejor se ajuste a sus datos.

Dado el concepto de proceso estacionario anteriormente definido, los modelos de pronóstico se dividen en:

❖ Modelos lineales estacionarios

- a. **Modelos autorregresivos:** “Los modelos autorregresivos (AR) expresan Y_t como una función lineal de cierto número de valores anteriores reales de Y_t ” (Hanke J,E.1996).
- b. **Modelo AR(p)**

Un modelo autorregresivo de orden p , o abreviadamente un modelo AR(p), se define de la siguiente forma:

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

Donde:

Y_t : Variable dependiente.

$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$: Variables independientes que son variables dependientes desfasadas un número específico de periodos.

$C, \phi_1, \phi_2, \dots, \phi_p$: Coeficientes de regresión.

ε_t : Es el término de residuo que representa sucesos aleatorios no explicados por el modelo. También se le conoce como “ruido blanco”, y $\varepsilon_t \sim (0, \sigma^2)$.

Usando el operador polinomial de retardos L, se denota como:

$$Y_t = C + \phi_1 LY_t + \phi_2 L^2 Y_t + \dots + \phi_p L^p Y_t + \varepsilon_t$$

Realizando algunas transformaciones tenemos:

$$Y_t - \phi_1 LY_t - \phi_2 L^2 Y_t - \dots - \phi_p L^p Y_t = C + \varepsilon_t$$

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) Y_t = C + \varepsilon_t$$

La expresión entre paréntesis como un polinomio en el operador de retardos L, se puede expresar de forma compacta:

$$\phi(L) Y_t = C + \varepsilon_t$$

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

Para que el proceso sea estacionario se requiere que las raíces de la ecuación polinomial estén fuera del círculo unidad.

$$1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p = 0$$

- **Modelo AR(1):** El caso más sencillo corresponde a un modelo autorregresivo de primer orden, donde el parámetro C es igual a cero.

El modelo autorregresivo de primer orden, viene definido por:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t, \quad |\phi_1| < 1$$

Utilizando el operador de retardos, por $(1 - \phi_1 L)Y_t = \varepsilon_t$

Cada variable ruido blanco influye sobre los valores de Y correspondientes al mismo periodo, o a periodos posteriores, pero nunca ejerce influencia sobre los valores de Y correspondientes a periodos anteriores.

Una consecuencia importante es que $E[\varepsilon_t Y_{t-\tau}] = 0$,

$$\forall \tau > 0$$

Para que el proceso AR(1) definido sea estacionario, la raíz del polinomio característico $1 - \phi_1 L = 0$, debe caer fuera del círculo unidad. Es decir:

$$|L| = |1/\phi_1| > 1, \text{ lo que equivale a que } |\phi_1| < 1$$

- **Modelo AR(2):** Un modelo autorregresivo de segundo orden, viene definido por:

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

Utilizando el operador de retardos,

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

Para que el proceso anterior sea estacionario se requiere que las raíces de la ecuación estén situadas fuera del círculo unidad, es decir:

$$1 - \phi_1 L - \phi_2 L^2 = 0$$

Si se cumplen las condiciones de estacionariedad se verificará que

$$E(Y_t) = 0$$

Condiciones de estacionariedad

Para que un modelo AR de orden p,

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t$$

Sea estacionario, se debe cumplir lo siguiente:

Usando el operador polinomial de retardos L:

$$Y_t = \varphi_1 L Y_t + \varphi_2 L^2 Y_t + \dots + \varphi_p L^p Y_t + \varepsilon_t$$

Realizando algunas transformaciones se tiene:

$$Y_t - \varphi_1 L Y_t - \varphi_2 L^2 Y_t - \dots - \varphi_p L^p Y_t = \varepsilon_t$$

$$(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) Y_t = \varepsilon_t$$

Y en forma compacta: $\phi(L) Y_t = \phi_p(L) Y_t = \varepsilon_t$

Para que el proceso sea estacionario se requiere que las raíces de la ecuación polinomial estén fuera del círculo unidad.

$$1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p = 0$$

La estacionalidad de la serie Y_t requiere, entre otras condiciones una media que no varía; es decir no debe existir una tendencia a lo largo del tiempo.

El polinomio autorregresivo usando el operador de retardo L , es:

$$\varphi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p$$

➤ **Modelos de medias móviles**

“Los modelos autorregresivos (AR) expresan Y_t como una función lineal de cierto número de valores anteriores reales de Y_t , mientras que los modelos de promedio móvil (MA) proporcionan pronósticos de Y_t con base en una combinación lineal de errores anteriores de Y_t ” (Hanke J, E.1996)

“Un modelo de medias móviles explica el valor de una determinada variable en un período t en función de un término independiente y una sucesión de errores correspondientes a períodos precedentes, ponderados convenientemente” (Uriel, E.1985).

Se denotan normalmente con las siglas MA, como en el caso de los modelos autorregresivos, ya indicado. Así, un modelo con q términos de error se denota como MA(q).

➤ **Modelo MA(q)**

Un modelo de medias móviles de orden q , MA(q), se define de la siguiente forma:

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

Donde:

Y_t : Variable dependiente.

$\theta_1, \theta_2, \dots, \theta_q$: Peso específico.

ε_t : Residuo o error. También conocido como “ruido blanco”.

$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$: Valores previos de residuos.

Utilizando el operador polinomial de retardos,

$$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$$

El modelo de medias móviles se puede expresar de forma compacta

$$Y_t = \theta(L)\varepsilon_t$$

Donde la media es cero, cualesquiera que sean los valores de θ_i es decir:

$$E[Y_t] = \theta(L)E(\varepsilon_t) = 0$$

Si en el modelo **MA (q)** se incluye un término constante

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

$$E[Y_t] = \mu$$

Donde:

Y_t : Variable dependiente.

$\mu, \theta_1, \theta_2, \dots, \theta_q$: Constante y peso específico.

ε_t : Error aleatorio o residuo (ruido blanco).

$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$: Valores previos de residuos.

Así pues, en los modelos de medias móviles, la media del proceso coincide con el término independiente, que aparece en el segundo miembro. Sin pérdida de generalidad se supondrá en lo sucesivo que $\mu = 0$.

Para que un proceso MA(q) sea invertible se requiere que las raíces de la ecuación polinomial caigan fuera del círculo unidad.

$$1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q = 0$$

Calculando los momentos del proceso, a partir del operador de retardo L se tiene:

$$\begin{aligned} 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q &= 0 \\ E(Y_t) &= E(\mu + \varepsilon_t - \theta_1 L \varepsilon_t - \theta_2 L^2 \varepsilon_t - \dots - \theta_q L^q \varepsilon_t) \\ &= E(\mu) + E(\Theta_q(L) \varepsilon_t) = \mu + \Theta_q(L) E(\varepsilon_t) \end{aligned}$$

Respecto de la varianza, se tiene:

$$V(Y_t) = E(Y_t^2) = E\left[\left(\varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}\right)^2\right]$$

Modelo MA (1)

Un modelo MA (1) viene definido por:

$$Y_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1}$$

Donde ε_t es un ruido blanco, un modelo MA(1) omitiendo la constante viene definido por:

$$Y_t = \varepsilon_t - \theta_1\varepsilon_{t-1}$$

$$Y_t = (1 - \theta_1L)\varepsilon_t$$

Para que la ecuación sea estable se requiere que la raíz del polinomio característico caiga fuera del círculo unidad, es decir

$$1 - \theta_1L = 0$$

$$|L| = \left|\frac{1}{\theta_1}\right| > 1; \theta_1 \neq 0$$

O, de forma equivalente, que $|\theta_1| < 1$

La condición de invertibilidad de un modelo MA(1) es equivalente en sentido formal a la condición de estacionariedad de un modelo AR(1). Un modelo MA(1) es siempre estacionario y la condición de invertibilidad se establece para poder pasar a un modelo AR(∞).

Modelo MA (2)

Un modelo MA (2) omitiendo la constante viene definido por:

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

Donde ε_t es un ruido blanco.

Para que un proceso MA (2) sea invertible se requiere que las raíces del polinomio característico caigan fuera del círculo unidad.

$$1 - \theta_1 L - \theta_2 L^2 = 0$$

Condiciones de estacionariedad e invertibilidad en los procesos MA.

Los modelos de medias móviles finitos son siempre estacionarios. Como puede apreciarse de las deducciones anteriores, los momentos de los procesos (esperanza, varianza y autocovarianzas) son invariantes en el tiempo. A diferencia de los modelos autorregresivos (AR), para la deducción de los momentos no es necesario suponer la estacionariedad de la serie. Si las raíces del polinomio de medias móviles caen fuera del círculo unidad, el proceso será también invertible.

Veamos el caso de un proceso MA(1) (omitiendo la constante para simplificar la exposición). Se tiene:

$$\begin{aligned}\varepsilon_t &= Y_t + \theta_1 \varepsilon_{t-1} \\ \varepsilon_{t-1} &= Y_{t-1} + \theta_1 \varepsilon_{t-2}\end{aligned}$$

Sustituyendo la segunda expresión en la primera:

$$\varepsilon_t = Y_t + \theta_1 [y_{t-1} + \theta_1 \varepsilon_{t-2}]$$

$$\varepsilon_t = Y_t + \theta_1 y_{t-1} + \theta_1^2 \varepsilon_{t-2}$$

Y continuando con la sustitución recursiva:

$$\begin{aligned} \varepsilon_t &= Y_t + \theta_1 Y_{t-1} + \theta_1^2 Y_{t-2} + \dots + \theta_1^i Y_{t-i} + \dots \\ Y_t &= \varepsilon_t - (\theta_1 Y_{t-1} + \theta_1^2 Y_{t-2} + \dots + \theta_1^i Y_{t-i} + \dots) \\ \Rightarrow Y_t &= -\sum_{i=1}^{+\infty} \theta_1^i Y_{t-i} + \varepsilon_t \end{aligned}$$

Alternativamente:

$$\begin{aligned} \varepsilon_t &= Y_t + \theta_1 Y_{t-1} + \theta_1^2 Y_{t-2} + \dots + \theta_1^i Y_{t-i} + \dots \\ \varepsilon_t &= Y_t (1 + \theta_1 L + \theta_1^2 L^2 + \dots + \theta_1^i L^i + \dots) \end{aligned}$$

Es decir, en principio es posible expresar el modelo MA(1) como un AR(∞).

Para que efectivamente ambos modelos sean equivalentes se requiere que el modelo AR sea estacionario, lo que impone la condición que:

$$|\theta| < 1$$

De esta forma, cuando el polinomio de medias móviles $\theta_q(L)$ (L) tiene sus raíces fuera del círculo unidad, el proceso de medias móviles puede transformarse en un proceso AR estacionario.

Otra forma de mostrar la invertibilidad para procesos MA (1).

$$\begin{aligned} Y_t &= \varepsilon_t - \theta_1 \varepsilon_{t-1} = (1 - \theta_1 L) \varepsilon_t \\ \Rightarrow \varepsilon_t &= \frac{1}{(1 - \theta_1 L)} Y_t \end{aligned}$$

Modelos mixtos autorregresivos de medias móviles (ARMA)

La combinación de modelos Autorregresivos (AR) y de Medias Móviles (MA) da lugar al modelo ARMA. Un modelo ARMA (p,q) se define de la siguiente forma:

$$Y_t = C + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

donde:

ε_t : es un ruido blanco.

φ_i : Constantes de regresión

θ_i : Pesos específicos

Utilizando los operadores polinomiales de retardo, el modelo se expresará de forma compacta de la siguiente forma:

$$\theta(L)Y_t = \theta(L)\varepsilon_t$$

➤ Para que el modelo sea estacionario se requiere que las raíces de la ecuación polinomial caigan fuera del círculo unidad.

$$\theta(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p = 0$$

Si se cumplen las condiciones de estacionariedad, el modelo ARMA (p,q) se puede expresar como un $MA(\infty)$, pudiendo representarse de la siguiente forma:

$$Y_t = \frac{\theta(L)}{\phi(L)} \varepsilon_t = \psi(L)\varepsilon_t$$

Por tanto los coeficientes del operador polinomial $\psi(L)$, que tiene infinitos elementos, deben cumplir la siguiente identidad:

$$\phi(L)\psi(L) = \theta(L)$$

En notación más detallada, se establece la siguiente identidad:

$$(1 - \phi_1 L - \dots - \phi_p L^p)(1 + \psi_1 L + \psi_2 L^2 + \dots) = (1 - \theta_1 L - \dots - \theta_q L^q)$$

A partir de la anterior identidad, se pueden deducir un conjunto de ecuaciones

que nos permiten obtener los ψ_i , en función de los coeficientes ϕ_h y θ_j .

Así en un modelo ARMA (1,1), la identidad anterior sería:

$$(1 - \phi_1 L)(1 + \psi_1 L + \psi_2 L^2 + \dots) = (1 - \theta_1 L)$$

- Para que un modelo ARMA (p,q) sea invertible, se requiere que las raíces de la ecuación polinomial caigan fuera del círculo unidad.

$$\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q = 0$$

Modelos ARMA (1,1)

Un proceso ARMA (1,1) (excluyendo la constante) se define:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

El proceso ARMA (1,1) es estacionario cuando $|\phi| < 1$, e invertible cuando

$$|\theta| < 1$$

Multiplicando ambos miembros por Y_{t-k} y tomando esperanzas, tenemos:

$$Y_k = E[Y_t Y_{t-k}]$$

$$Y_k = \phi_1 Y_{k-1} + E[\varepsilon_t Y_{t-k}] - \phi_1 E[\varepsilon_{t-1} Y_{t-k}]$$

Teniendo en cuenta que:

$$E[\varepsilon_t Y_t] = \sigma^2$$

$$E[\varepsilon_{t-1} Y_t] = E[\varepsilon_{t-1} (\phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1})] = (\phi_1 - \theta_1) \sigma^2$$

La expresión se deduce a las siguientes expresiones:

$$\text{Para } k=0 \rightarrow Y_0 = \phi_1 Y_1 + \sigma^2 - \theta_1 (\phi_1 - \theta_1) \sigma^2$$

$$\text{Para } k=1 \rightarrow Y_1 = \phi_1 Y_0 - \theta_1 \sigma^2$$

Sustituyendo este valor en la expresión de la varianza (Y_0) se tiene:

$$Y_0 = \phi_1 (\phi_1 Y_0 - \theta_1 \sigma^2) + \sigma^2 - \theta_1 (\phi_1 - \theta_1) \sigma^2$$

$$Y_0 = \frac{1 - 2\theta_1 \phi_1 + \theta_1^2}{1 - \phi_1^2} \sigma^2$$

Que vuelve a sustituirse en la expresión de la covarianza de primer orden (Y_1)

$$\text{Para } k > 1 \rightarrow Y_k = \phi_1 Y_{k-1}$$

De esta forma, los coeficientes de autocorrelación quedan como:

$$P_k = \begin{cases} \frac{(1-\varphi_1\theta_1)(\varphi_1\theta_1)}{1-2\theta_1\varphi_1+\theta_1^2}; & k=1 \\ \varphi_1 P_{k-1} & ; k > 1 \end{cases}$$

Es decir, los coeficientes de autocorrelación de un ARMA(1,q) se comportan como un AR(1) puro para $k > 1$.

❖ Modelos lineales no estacionales

Modelos de promedio móvil autorregresivo integrado: ARIMA (p, d, q).

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

“Los modelos de promedio móvil autorregresivo integrado (ARIMA) son una clase especializada de técnicas de filtración que ignoran por completo a las variables independientes en la formulación de pronósticos. Estos modelos son dispositivos altamente refinados de ajuste de curvas que utilizan valores reales y anteriores de la variable dependiente, para producir pronósticos precisos de corto plazo” (Hanke J, E.1996)

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. Podemos decir que la consideración exclusiva de los valores pasados de una determinada variable para explicar su evolución presente y futura supone, al mismo tiempo una ventaja y un inconveniente:

La ventaja radica en el hecho de no necesitar distintas series de datos (distintas variables) referida al mismo periodo de tiempo (característica común a todos los

modelos univariantes) y al mismo tiempo, ahorramos la identificación y especificación del modelo en el sentido de la econometría tradicional.

El inconveniente es que, al renunciar a la inclusión de un conjunto más amplio de variables explicativas, no atendemos a las relaciones que sin duda existen entre casi todas las variables económicas perdiendo capacidad de análisis de tiempo que renunciamos, implícitamente, al estudio teórico previo del fenómeno y a su indudable utilidad.

Los modelos ARIMA (p,d,q) con p: número específico de periodos desfasados, q: número de valores previos y d: número de diferenciaciones; constituyen una clase particular de procesos no estacionarios y se define formalmente como:

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

Donde:

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

ϕ_i : número de retardos para la variable Y

ω_i : número de valores previos de residuo ε_t

ε_t : error aleatorio (ruido blanco)

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

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 económicas y, al mismo tiempo, posibles de ser transformados en procesos estacionarios. En primer lugar, analizaremos el proceso de "caminata aleatoria".

Función de autocorrelación (FAC)

Se usa para ver si la serie es estacionaria o no estacionaria. La función de autocorrelación mide la correlación entre los valores de la serie distanciados un lapso de tiempo k .

Dada una muestra Y_0, Y_1, \dots, Y_{n-1} de n observaciones, la función de autocorrelación muestral (FAC) de la muestra al rezago k , denotada por ρ_k , se define como:

$$\rho_k = \begin{cases} \gamma_k & \text{para } k = 1, \dots, n-1 \\ 1 & \text{para } k = 0 \end{cases}$$

Donde:

γ_0 : es la varianza de la muestra dada por:

$$\gamma_0 = \frac{1}{n-1} \sum_{t=1}^n (Y_t - \bar{Y})^2$$

$$\gamma_k = \frac{1}{n} \sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})$$

“La función de autocorrelación indica cuánta correlación existe entre datos individuales contiguos en la serie Y_t . Conforme el valor del retraso aumenta, el número de observaciones comprendidas en la autocovarianza disminuye hasta el elemento final” (Pindick R, S. & Rubinfeld D, L.2003). Al graficar ρ_k frente a k , la gráfica obtenida se conoce como correlograma.

Función de autocorrelación parcial (FACP)

“Se emplea para ayudar a identificar el grado de relación entre los valores reales de una variable y valores anteriores de la misma, mientras que se mantienen constantes los efectos de las otras variables (periodos retrasados)” (Hanke J, E. 1996).

“La función de autocorrelación parcial (FACP) de la muestra ρ_{kk} en el retraso k es la correlación entre observaciones (series de tiempo) que están separadas k periodos de tiempo, manteniendo constantes las correlaciones en los rezagos intermedios (es decir rezagos menores de k). En otras palabras, la autocorrelación parcial es la correlación entre Y_t y Y_{t-k} después de eliminar el efecto de las Y intermedias.

En la siguiente tabla 1 se dan algunos lineamientos generales acerca de los patrones típicos de las funciones de correlaciones muestral y parcial (Gujarati N., 2004).

Tabla 1:
Patrones teóricos de FAC y FACP

Tipo de modelo	Patrón típico de FAC	Patrón típico de FACP
AR(p)	Disminuye exponencialmente o con un patrón sinusoidal decreciente o ambos	Picos grandes a lo largo de los p rezagos
MA(q)	Picos grandes a lo largo de los q rezagos	Decrece exponencialmente
ARMA(p, q)	Decrece exponencialmente	Decrece exponencialmente

Fuente: (Gujarati, 2004)

Caminata aleatoria

Es una serie de tiempo estocástica en la que cada cambio sucesivo en Y_t , expresado como u_t es extraído en forma independiente de una distribución de probabilidad con media 0 y varianza σ^2 . Por lo tanto, Y_t está determinada por:

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

Caso general

Dada una serie Y_t , que eventualmente corresponde a los logaritmos de los valores originales, si su diferencia de orden "d" puede ser representada por un proceso ARMA(p,q) estacionario, se dice que la serie Y_t sigue un proceso ARIMA(p,d,q).

La letra "I" en ARIMA corresponde a "Integración", la operación inversa a la diferenciación.

Si $Z_t = \Delta^d Y_t$ y Z_t sigue un proceso ARMA(p,q) estacionarios:

$$(1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p) Z_t = (1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q) \varepsilon_t$$

$$\varphi_p(L) Z_t = \theta_q(L) \varepsilon_t$$

Entonces Y_t sigue un proceso ARIMA (p,d,q). También se escribe en la variable original Y_t como:

$$\phi_p(L)(1-L)^d Y_t = \theta_q(L) \varepsilon_t$$

Generalmente, no son necesarias diferencias regulares de orden superior a 2, excepto en el caso de variables que presentan estacionalidad.

Transformación Box - Cox: “una transformación instantánea en el sentido de que no están involucrados simultáneamente varios periodos de tiempo de carácter más general que la transformación logarítmica”(Box G, E. P. & Cox, D. R.1964). Esta transformación se define por:

$$Y_t^\lambda = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda}; \lambda \neq 0 \\ \text{Ln}(Y_t); \lambda = 0 \end{cases}$$

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

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

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

Modelos ARIMA estacionales

Un modelo estacional puro se caracteriza porque sólo existe relación entre las observaciones que distan entre sí S periodos o múltiplos de S . Son series con

ciclos u oscilaciones estrictamente periódicas, donde el periodo es igual o inferior al año. El periodo estacional se designa por “S”, así en datos trimestrales s=4, en datos anuales s=12, etc. La elaboración de modelos ARIMA estacionales presentan características análogas a los modelos ARIMA no estacionales.

Los métodos que emplean modelos estacionales ó SARIMA(p,d,q)*(P,D,Q) suponen que el componente estacional es generado por un proceso estocástico, cuya identificación se realiza de manera similar a los modelos que representan la estructura regular de una serie, con la excepción de que para ello se examinan los “valores estacionales” de las funciones de autocorrelación (valores que corresponden a los rezagos 4,8,12, ... si los datos son trimestrales y 12, 24, 36, ... si los datos son mensuales). De este modo una serie podría requerir diferencias de orden estacional si los valores estacionales de la función de autocorrelación no tienden a cero rápidamente.

Un proceso SARIMA (p, d, q)*(P,D,Q) se define así,

$$\phi_p(L)\Phi_P(L^s)(1-L)^d(1-L^s)^D Y_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t$$

Donde:

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

$$\theta_q(L) = 1 - \theta_1 L - \theta_2 L^2 - \dots - \theta_q L^q$$

$$\Phi_P(L^s) = 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_P L^{Ps}$$

$$\Theta_Q(L^s) = 1 - \Theta_1 L^s - \Theta_2 L^{2s} - \dots - \Theta_Q L^{Qs}$$

Intérvalos de confianza para las predicciones

La varianza del error de predicción puede utilizarse para obtener intervalos de confianza de las predicciones elaboradas, mediante la expresión:

$$\Pr \left[\tilde{y}_{T+k} \pm \lambda_{\alpha} \widehat{\sigma}_{e_T(k)} \right]$$

Donde, si se supone que la innovación ε_t sigue una distribución normal, el parámetro λ_{α} se obtendrá de las tablas de dicha distribución, al nivel de confianza α elegido.

Medición del error de pronóstico

Puesto que las técnicas cuantitativas de pronóstico con frecuencia implican datos de series de tiempo, se desarrolló una notación matemática para referirse a cada periodo específico. La notación básica para pronósticos se resume de la siguiente forma:

Y_t : valor de una serie de tiempo en el periodo t.

\widehat{Y}_t : valor pronosticado de Y_t .

$e_t = Y_t - \widehat{Y}_t$: residual o error residual de pronósticos.

Un método para calcular las técnicas de pronósticos utiliza la suma de errores absolutos. El error absoluto medio (MAE), mide la precisión del pronóstico al promediar las magnitudes de los errores de pronóstico (valores absolutos de cada error). MAE es más útil cuando el analista quiere medir el error de pronóstico en las

mismas unidades que la serie original. La siguiente ecuación muestra cómo se calcula el MAE.

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |Y_t - \hat{Y}_t|$$

El error cuadrático medio (MSE) es otro método para evaluar una técnica de pronóstico. Cada error de pronóstico o residual se eleva al cuadrado; luego, se suman y se dividen entre el número de observaciones. Este método penaliza los errores grandes de pronóstico debido a que los errores se elevan al cuadrado, lo cual es importante; una técnica que produce errores moderados podría ser preferible a una que, por lo general, tiene errores pequeños, pero que en ocasiones produce errores muy grandes (Hanke, J.E. & Wichern, D.W. 2006). El error cuadrático medio (MSE) está dado por la siguiente ecuación:

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2$$

A veces, es más útil calcular los errores en términos de porcentajes en lugar de cantidades. El error porcentual absoluto medio (MAPE) se calcula al encontrar el error absoluto en cada periodo, dividiéndolo entre el valor real observado para ese periodo y luego promediando los errores porcentuales absolutos. El MAPE proporciona una indicación de cuán grandes son los errores de pronóstico en comparación con los valores reales de la serie. Esta técnica es especialmente útil cuando valores de Y_t son grandes. También puede utilizarse el MAPE para comparar la precisión de las mismas o diferentes técnicas en dos series totalmente distintas. La siguiente ecuación muestra cómo se calcula el MAPE.

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t}$$

El criterio de la información de Akaike o ajuste del modelo y el número de parámetros (AIC)

Es seleccionar el mejor modelo de un grupo de modelos candidatos como aquel que minimiza es:

$$AIC = Ln\hat{\sigma}^2 + \frac{2}{n} r$$

$\hat{\sigma}^2$: La suma de cuadrados de los residuales dividida entre el número de las observaciones

n : El número de las observaciones (residuales)

r : El número total de los parámetros (mas el termino constante) en el modelo

ARIMA

Metodología de “Box-Jenkins”

Es uno de los métodos predictivos y se fundamenta en la estimación eficiente de los parámetros por medio de los procesos iterativos.

De acuerdo a Box-Jenkins (1970), el análisis de series de tiempo implica las siguientes etapas: (i) Identificación, (ii) Estimación, (iii) Verificación y (iv) Pronóstico (Predicción). Si la serie es débilmente estacionaria, se procede de inmediato con la etapa (i); caso contrario, la serie debe ser "pre-procesada" a fin de ser transformada en realizaciones estacionarias. Asumiendo que se cuenta con series estacionarias,

la identificación tiene por objeto determinar el tipo de modelo a aplicar (AR, MA ó ARMA) y el orden de los parámetros "p" y "q".

Un aspecto importante en la modelación ARIMA de una serie de tiempo simple es el número de veces que ésta necesita de una diferencia antes de fijar el modelo.

2.2.2. Redes Neuronales.

- Los primeros trabajos en redes neuronales fueron realizados por (McCulloch W, S. & Pitts, W, H. 1943), donde buscaban la representación computacional de una neurona biológica. Las redes neuronales proveen un modelo no lineal cuya estructura se dice: "imita la estructura del cerebro humano" de alguna manera. Una Red Neuronal se puede describir como un sistema interconectado de manera no lineal entre un conjunto de entradas con un conjunto de salidas. La idea de realizar predicciones en series temporales con redes neuronales se remonta alrededor de 1960, (Zhang & PatuwoyHu, 1998) donde investigadores y científicos empezaron a darse cuenta de todas las ventajas que podía otorgar una red neuronal para realizar la predicción, como es la capacidad de generalización y de poder aprender mediante la experiencia.
- Dado el vector aleatorio $X \in \mathbb{R}^p$ y la variable dependiente $Y \in \mathbb{R}$, una red neuronal es un modelo estadístico no lineal, definido con la relación:

$$Y = f(X) + \varepsilon = g^s \left[\beta_{ok} + \beta_k^T \sigma(w_{om} + w_m^T X) \right] + \varepsilon$$

Donde

$g^s = g.g.g\dots g$, se construye de forma recursiva (composición de la función g , s -veces) a partir de aproximaciones no-lineales.

g^s , σ son funciones de activación o transferencia.

β , w son vectores de pesos.

s : es el número de capas ocultas.

Frecuentemente en muchos problemas, la red de una sola capa es suficiente para describir el comportamiento de los datos. El modelo estadístico no lineal para este caso, está definido por:

$$Y = f(X) + \varepsilon = g^s \left[\beta_{ok} + \beta_k^T \sigma(w_{om} + w_m^T X) \right] + \varepsilon , \text{ cuando } s = 1$$

2.2.2.1. Ecuaciones matemáticas con el modelo ARIMA y SARIMA

Al respecto se toma como referencia a (Hyndman J, R. y Khandakar, Y. 2008), por lo que para obtener las ecuaciones matemáticas se parte de la siguiente expresión general:

$$\underbrace{\left(1 - \sum_{i=1}^p \phi_p L^i \right)}_{AR(p)} \underbrace{\left(1 - \sum_{i=1}^P \Phi_P L^{Ps} \right)}_{AR(P)} \underbrace{\left(\prod_{i=1}^d (1-L)^d \right)}_d \underbrace{\left(\prod_{i=1}^D (1-L^S)^D \right)}_D Y_t = \underbrace{\left(1 + \sum_{i=1}^q \theta_q L^i \right)}_{MA(q)} \underbrace{\left(1 + \sum_{i=1}^Q \Theta_Q L^{Qs} \right)}_{MA(Q)} \varepsilon_t$$

Donde:

L : operador de desplazamiento hacia atrás

Y_t : Variable a determinar

p : Orden de la parte autorregresiva no estacional

d : Grado de diferenciación no estacional

q : Orden de la parte de la media móvil no estacional

P : Orden de la parte autorregresiva estacional

D : Grado de la diferenciación estacional

Q : Orden de la parte de la media móvil estacional

ϕ_i : Coeficiente de la parte autorregresiva no estacional

θ_i : Coeficiente de la parte de la media móvil no estacional

Φ_i : Coeficiente de la parte autorregresiva estacional

Θ_i : Coeficiente de la parte de la media móvil estacional

S : Grado de la diferenciación estacional

ε_i : Parte aleatoria

2.2.2.2. Criterio de información Akaike (AIC)

El AIC se expresa mediante la ecuación matemática siguiente:

$$AIC = n + n \times \ln(2\pi) + n \times \ln\left(\frac{SSE}{n}\right) + 2 \times (k + 1)$$

Donde n : es número de datos observados, k : número de parámetros y SSE : es la suma de cuadrados de los errores

2.2.3. Neuronas Biológicas.

Las Redes Neuronales Artificiales persiguen imitar ciertas habilidades humanas atribuibles al cerebro y a millones de elementos interconectados llamados Neuronas. Las neuronas son células nerviosas que constituyen los elementos primordiales del sistema nervioso central. Son capaces de recibir señales

provenientes de otras neuronas, procesar estas señales, generar pulsos nerviosos, conducir estos pulsos, y transmitirlos a otras neuronas. (Colina,E. & Rivas, F. 1998).

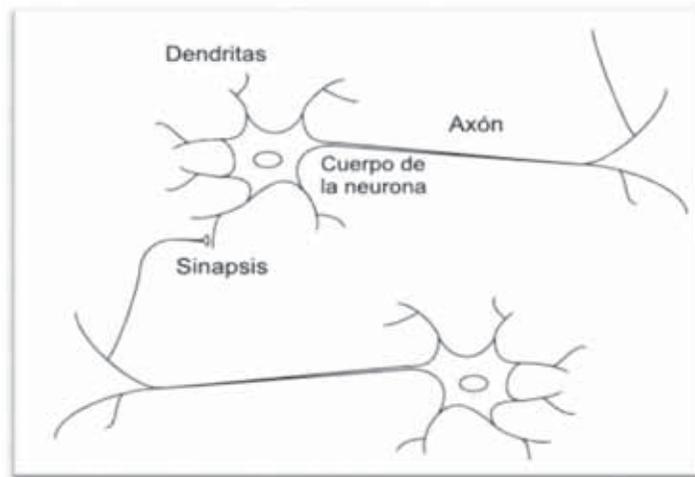


Figura 1: Esquemático de Neuronas Biológicas.

“Las dendritas son como un árbol de redes receptoras de fibras nerviosas que llevan señales eléctricas al cuerpo de la célula” (Hagan M, T., Dermuth H, B. & Beale, M.1996).

“El cuerpo celular contiene el núcleo, tiene forma esférica y es aquí donde se ejecutan todas las transformaciones necesarias para la vida de la neurona” (Colina, E. & Rivas, F.1998).

El axón transmite la señal de salida a otras neuronas. El intercambio químico de información entre una neurona y otra se hace a través de la sinapsis. Ésta es el punto de interconexión entre neuronas.

Modelo de una Neurona Artificial

“Una neurona artificial o nodo es una unidad de procesamiento de información que es fundamental para la operación de una red neuronal” (Haykin, S.1994). En un modelo de una neurona pueden identificarse los siguientes elementos:

Entradas o nodos de entrada: son escalares que le proporcionan a la red, de acuerdo al problema de estudio.

- a. Salidas o nodos de salida: son los valores que origina la red como resultado del aprendizaje.
- b. Un conjunto de pesos sinápticos o simplemente pesos: son valores numéricos que expresa la importancia de la entrada correspondiente. El valor de la entrada X_i se dirige a la neurona k multiplicado por el W_{ik} (el primer subíndice de los pesos se refiere a la entrada y el segundo a la neurona en cuestión).
- c. Un punto de suma de entradas ponderadas: aquí se realiza la combinación lineal o suma de todas las entradas multiplicadas por sus correspondientes pesos.
- d. Una función de activación: es una función, que puede ser lineal o no lineal, que limita el rango de la salida de la neurona.
- e. Sesgo: es un valor formado por una entrada fija e igual a 1, denominada X_0 , multiplicada por el peso W_{0k} .

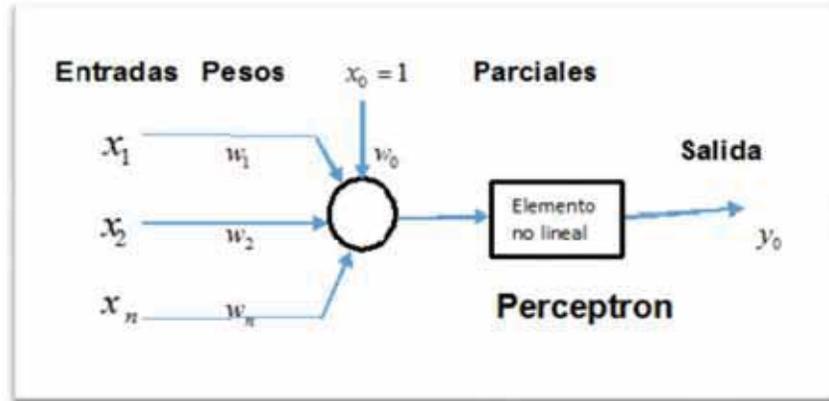


Figura 2: Modelo de una Neurona.

En la figura 2 se observa que la neurona k puede describirse mediante las siguientes ecuaciones:

a. **La acumulación,**

$$v_k = \sum_{i=0}^p x_i w_{ik}$$

b. **La función de activación,**

$$y_k = \Gamma(v_k)$$

Donde: x_0, x_1, \dots, x_p son las entradas de la neurona k ; $w_{0k}, w_{1k}, \dots, w_{pk}$

son los pesos de la neurona k ; v_k es la suma de las entradas multiplicadas por los

pesos correspondientes; $\Gamma(\cdot)$ es la función de activación i y_k es la salida de la neurona k .

Los pesos son parámetros escalares que se van ajustando según la aplicación de una regla de aprendizaje de manera de cumplir con la relación entrada/salida.

La función de activación se selecciona de acuerdo al objetivo del problema y al rango de valores en que se requiera la salida.

Funciones de Activación

Existen diversos tipos de funciones de activación, entre los más utilizados están:

a.- Función Paso: La salida de este tipo de función de activación pueden ser 0 o 1, dependiendo si el parámetro de la función es positiva o negativo. Se usa para problemas de clasificación. Su expresión matemática y gráfica es:

$$\Gamma(v) = \begin{cases} 1, & \text{si } v \geq 0 \\ 0, & \text{si } v < 0 \end{cases}$$

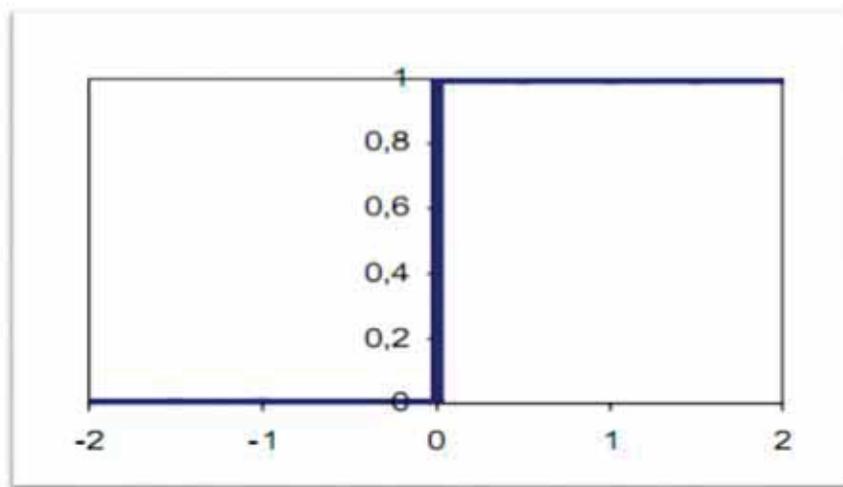


Figura 3: Función Paso

b.- Función Lineal: La entrada de la función de activación es igual a la salida. Se usa en diversos tipos de redes, con frecuencia, en la capa de salida. Su expresión matemática y gráfica es:

$$\Gamma(v) = v$$

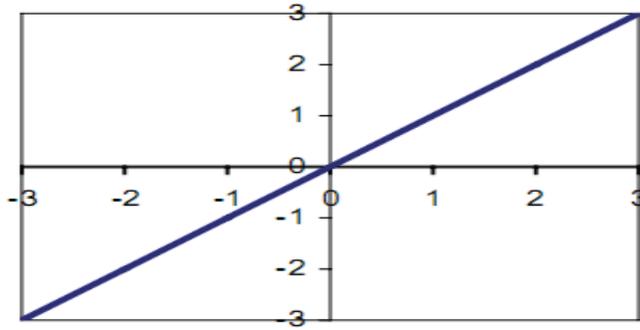


Figura 4: Función Lineal.

c.- Función Rampa: Su salida está entre -1 y 1. Su expresión matemática y gráfica es:

$$\Gamma(v) = \begin{cases} -1, & \text{si } v \leq -1 \\ v, & \text{si } -1 < v < 1 \\ 1, & \text{si } v \geq 1 \end{cases}$$

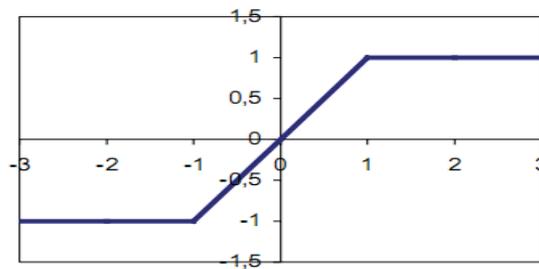


Figura 5: Función Rampa.

d.- Función Logística: Su salida comprende valores entre 0 y 1. Es la función de activación más usada en redes neuronales y se recomienda para problemas de predicción. Su expresión matemática y gráfica es:

$$\Gamma(v) = \frac{1}{1 + e^{-av}}$$

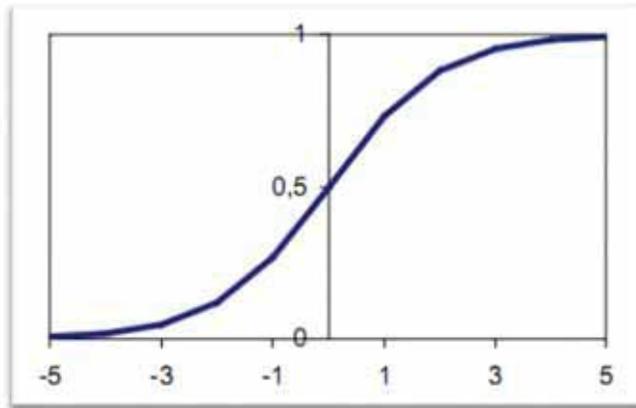


Figura 6: Función Logística.

Propiedades:

La función logística, modeliza el crecimiento logístico de una magnitud $M(t)$ tiene las siguientes propiedades:

i.- **crecimiento acotado**, $\forall t$ se cumple que:

$$0 < M(t) < K$$

ii.- **aproximación exponencial**, para valores pequeños de M/K (o también para valores de $t \rightarrow -\infty$ la función logística puede aproximarse por un modelo de crecimiento exponencial del tipo $M(t) \approx M' e^{rt}$

iii.- **valores límite**, la función logística generalizada dada por la ecuación, tiene los siguientes límites:

$$P(t; a, m, n, \tau) = a \frac{1 + m e^{-t/\tau}}{1 + n e^{-t/\tau}}$$

$$\lim_{t \rightarrow -\infty} P(t) = a \frac{m}{n}; P(0) = a \frac{1+m}{1+n}; \lim_{t \rightarrow \infty} P(t) = a$$

iv.- es continua y derivable,

$$y = \frac{1}{1 + e^{-x}}$$

$$y' = \frac{e^{-x}}{(1 + e^{-x})^2} = y(1 - y)$$

e.- Función Tangente Hiperbólica: Es semejante a la función logística, pero su salida está entre -1 y 1. Se utiliza con frecuencia en redes multicapa. Su expresión matemática y gráfica es:

$$\Gamma(v) = \text{Tanh}(v)$$

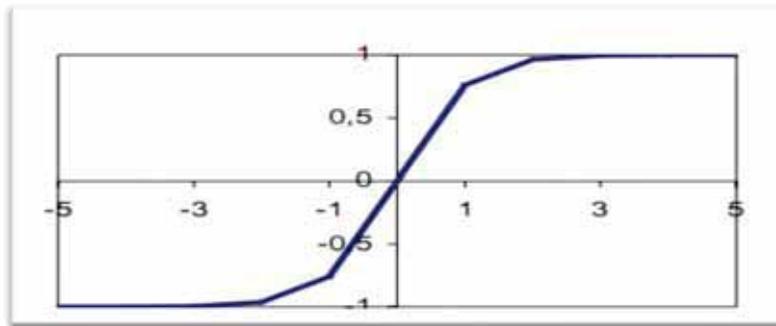


Figura 7: Función Tangente Hiperbólica.

f.- Función Gaussiana: Su rango está entre 0 y 1. Se utiliza en redes neuronales de función de base radial, las cuales pueden aplicarse a problemas de predicción. Su expresión matemática y gráfica es:

$$\Gamma(v) = e^{-v^2/e^2}$$

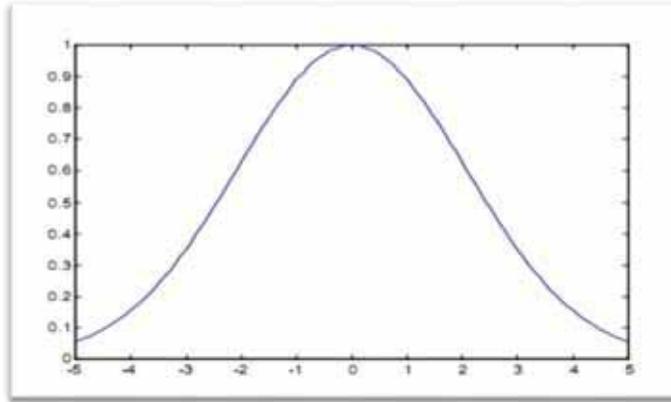


Figura 8: Función Gaussiana.

Arquitectura de Red

El término arquitectura de red se refiere a la forma o estructura de una red neuronal artificial (RNA). La arquitectura de un modelo de RNA incluye los siguientes componentes: entradas, capas y salidas; así mismo, la interconexión y dirección de la red, que se refiere a la forma en que se interrelacionan las neuronas. El algoritmo de aprendizaje está muy relacionado con la arquitectura de la red, pero se estudiará más adelante.

Componentes

- ✓ **Entradas:** Es el canal de alimentación de la red. Se deberán establecer el número de entradas de acuerdo al caso en estudio. Los datos de entradas son numéricos y en muchos casos pueden ser convenientes cambiar de escala y/o preprocesarlos. El escalamiento se refiere a un cambio de escala, convertirlos a datos entre 0 y 1, entre -1 y 1, o estandarizarlos, de acuerdo al rango de las funciones de activación involucradas. El preprocesamiento se refiere a la aplicación de algún método estadístico de (Box G & Cox, 1964) de exploración de datos que permita cualquier

transformación que mejore el conjunto original de datos en beneficio de un mejor desempeño de la red.

- ✓ **Capas:** También se les denomina capas ocultas o intermedias, por encontrarse entre la entrada y la salida de la red. Con frecuencia no es suficiente una sola neurona para resolver un problema, sino que se requiere de varias neuronas que operen en paralelo, lo que se denomina capa, e inclusive pudieran ser necesarias varias capas. A la red con una sola capa se le denomina red unicapa, mientras que a la red con dos o más capas se le denomina red multicapa.

El número de capas y de neuronas o nodos que componen cada capa debe especificarse en la arquitectura. El número de capas deberá ser mayor si el problema es no lineal y complejo, pero, en general, un problema podrá representarse bastante bien con una o dos capas ocultas. El número de neuronas por capa, puede variar entre una capa y otra. Se tiene algunos criterios para determinar el número de nodos por capa, éstos deberán determinarse por ensayo y error. Un criterio que pudiera ser útil es considerar al promedio entre el número de entradas y salidas como un valor referencial del número de nodos ocultos.

- ✓ **Salida:** Son neuronas o nodos de salida. El número de salidas de la red dependerá del problema de estudio. La salida de la red deberá estar expresada en la misma escala de los datos originales, es decir, si los datos fueron alterados, deberán ser regresados a su escala inicial. Si se realizó un pre procesamiento deberá hacerse un pos procesamiento.

Interconexión

Las interconexiones entre capas de las redes pueden clasificarse como:

- a. **Totalmente conectada:** La salida de una neurona de la capa i es entrada a todas las neuronas de la capa $i+1$.
- b. **Localmente conectada:** La salida de una neurona de la capa i es entrada a una región de neuronas de la capa $i+1$.

Dirección

La dirección de la información de las redes pueden clasificarse en:

- a. **Redes de alimentación adelantada:** Las salidas de las neuronas de una capa sólo se propagan a las neuronas de la capa siguiente. Es decir, la información fluye solamente de la entrada a la salida.
- b. **Redes retroalimentadas:** Las salidas de las neuronas de una capa pueden ser entradas de las neuronas de las capas anteriores.
- c. **Redes de alimentación lateral:** Las salidas de las neuronas pueden ser entradas de neuronas de la misma capa.
- d. **Redes recurrentes:** Existen lazos cerrados.

Un ejemplo de red unicapa totalmente conectada, de alimentación adelantada, con 3 entradas, 2 neuronas o nodos en la capa oculta, y una salida sería el siguiente

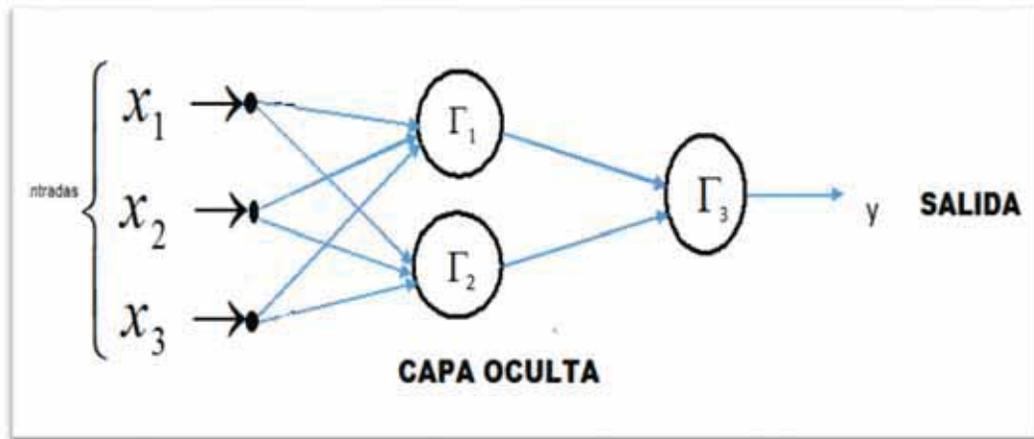


Figura 9: Ejemplo de una arquitectura de red.

Esta red, está formada por una sola capa que consta de dos neuronas en paralelo. “Las entradas se orientan a la salida. Las 3 entradas se comunican con las dos neuronas de la única capa, estas dos neuronas se comunican con la neurona de salida para proporcionar la salida de la red. No se consideran como capas las entradas ni las salidas” (Colina & Rivas, 1998) y (Hagan M, Dermuth H, & Beale, 1996).

Entrenamiento y Generalización

Un modelo de red neuronal al ser construido origina dos actividades fundamentales para la posterior utilización. Estas son:

Entrenamiento

“El entrenamiento de una red se lleva a cabo mediante una regla o algoritmo de aprendizaje. Este algoritmo es un procedimiento para modificar los pesos de una red y su propósito es entrenar la red para ejecutar alguna tarea”. (Hagan M, Dermuth H, & Beale, 1996), también (Colina E., & Rivas, F. 1998) “En otras palabras,

los algoritmos de aprendizaje proporcionan la habilidad para aprender a capturar información, mediante el ajuste de los pesos o ponderaciones de interconexiones”.

Terminología

En el proceso de entrenamiento se usa cierta terminología que conviene explicar previamente:

1. **Patrón:** Una observación o experimento del conjunto o patrones de entrenamiento.
2. **Patrones de Entrenamiento:** Son ejemplos o grupos de datos suministrados a la red para su aprendizaje y están compuesto por patrones de entrada y patrones de salida.
3. **Ciclo:** Es un barrido de todos los patrones del conjunto o patrones de entrenamiento.

Criterios de Paradas del Entrenamiento:

- a. Se repiten los ciclos hasta que el error alcance un valor máximo permisible.
- b. Se alcance un número máximo de ciclos.

Algoritmos de Entrenamiento

Los algoritmos de entrenamiento, en general, se pueden clasificarse en dos tipos: supervisados y no supervisados.

Supervisados: “Un maestro guía la red en cada etapa de aprendizaje, indicándole el resultado correcto” (Colina, E. & Rivas, F.1998). La misión del algoritmo es ajustar los pesos de la red de manera tal, que dado un conjunto de entradas las salidas proporcionadas por la red deberán coincidir lo más posible con las salidas

especificadas en el patrón de entrenamiento. Ejemplos: regla de Hebb, aprendizaje de Widrow-Hoff y algoritmos de retropropagación.

En esta investigación utilizaremos únicamente algoritmos de entrenamiento supervisados. Debido a su educación para resolver problemas de predicción. Son muchos los algoritmos de aprendizaje, pero se expondrán los que se consideran de interés para la comprensión de este punto y de este trabajo.

Perceptrón Discreto: Se utiliza para clasificar mediante una recta, es un dicotomizador. El algoritmo de aprendizaje para un perceptrón discreto con una sola neurona puede expresarse como sigue:

$$W_{(k+1)} = W_{(k)} + ceX \quad (2.1)$$

Donde:

$W_{(k)}$: es el vector de pesos para el patrón de entrenamiento k

$W_{(k+1)}$: es el vector de pesos para el patrón de entrenamiento k+1

c : Factor de aprendizaje

$e = Y_d - Y_n$: es el error, la diferencia entre la salida deseada (proveniente del patrón de entrada del conjunto de entrenamiento) y la salida estimada por la red neuronal.

X : Vector de entradas (proveniente del patrón de entrada del conjunto de entrenamiento)

Perceptrón Discreto Unicapa (múltiples neuronas): El algoritmo de perceptrón discreto, dado por la ecuación (2.1), actualizado el vector de una sola neurona. Este algoritmo puede generalizarse para el caso de múltiples neuronas como sigue:

$$W_{ij(k+1)} = W_{ij(k)} + ceX_i \quad (2.2)$$

Los vectores de pesos del perceptrón discreto ahora se convierten en matrices de pesos. Éstas contienen los pesos entre la entrada y las neuronas de la única capa y los pesos entre estas neuronas y la salida.

Perceptrón Continuo: La función de activación del perceptrón discreto se cambia en el perceptrón continuo a fin de lograr un control más fino del proceso de entrenamiento. Esta función deberá ser diferenciable, permitiendo de esa manera el cálculo de errores por gradiente. Para el ajuste de los pesos se minimiza la función del error utilizando procedimientos de máximo descenso o por técnicas de gradiente. El procedimiento básico de descenso consiste en calcular el gradiente de la función del error $\nabla E(W)$, empezando con un vector de pesos arbitrariamente seleccionado W ; el próximo valor de W es obtenido moviéndonos en la dirección negativa del gradiente sobre la superficie de error multidimensional. La dirección de gradiente negativo es aquella de descenso máximo. El algoritmo puede expresarse como:

$$W_{(k+1)} = W_{(k)} - \eta \nabla E(W_{(k)}) \quad (2.3)$$

Donde η es una constante positiva llamada factor de aprendizaje y $E = (Y_d - Y_n)^2$ es el error para el k-ésimo patrón de entrenamiento, “el cual consiste en la diferencia cuadrática entre la salida deseada y la salida de la red neuronal” (Colina, E. & Rivas, F.1998).

Perceptrón multicapa

El perceptrón multicapa es una red neuronal artificial formada por múltiples capas, esto le permite resolver problemas que no son linealmente separables. La salida de una neurona está determinada por la suma ponderada con las salidas de las neuronas de la capa anterior y una función de activación.

Soluciona la separabilidad en regiones convexas y precisa del algoritmo de retro propagación (Castillo, J. 2016).

Perceptrón multicapa para el pronóstico de series de tiempo

“Para modelar series de tiempo con comportamiento supuesto como lineal se han usado ampliamente modelos como: AR, MA, ARMA, y ARIMA” (Box, y Jenkins, 1994). Sin embargo, éste tipo de modelos no es suficiente, “dado que la gran mayoría de series de tiempo en ingeniería, finanzas y econometría presenta un comportamiento aparentemente no lineal” (Palit y Popovic, 2005).

Respecto al pronóstico de series de tiempo con RNA, Zhang *et al.* (2008) realizaron una revisión general del estado del arte donde resaltan tanto éxitos y fracasos reportados de las redes neuronales (especialmente con los perceptrones multicapa); incluyendo las publicaciones más relevantes y los tópicos de investigación más influyentes hasta 1996. Sin embargo, en la última década se ha producido un considerable número de contribuciones en múltiples campos como

metodologías de aprendizaje, selección de entradas relevantes, neuronas ocultas, entre otros, cuya influencia no ha sido evaluada ni reportada en la literatura.

Para un modelo de red neuronal artificial (RNA) con una capa oculta, **la ecuación general de predicción** para calcular un pronóstico de y_t (la salida) usando observaciones pasadas seleccionadas, $y_{t-j_1}, \dots, y_{t-j_k}$, como las entradas; es posible especificar matemáticamente como sigue:

$$\hat{y}_t = \phi_0 \left\{ w_{co} + \sum_h w_{ho} \phi_h \left(w_{ch} + \sum_i w_{ih} y_{t-j_i} \right) \right\}$$

Donde $\{w_{ch}\}$ denota los pesos para las conexiones entre la entrada constante y las neuronas ocultas y w_{co} denota el peso de la conexión directa entre la entrada constante y la salida. Los pesos $\{w_{ih}\}$ y $\{w_{ho}\}$ denotan los pesos para las otras conexiones entre las entradas y las neuronas ocultas, y entre estas neuronas y la salida, respectivamente. Las dos funciones ϕ_h y ϕ_o denotan las funciones de activación usadas en la capa oculta y en la salida respectivamente.

La notación RNA $(j_1, \dots, j_k; h)$ denota a la RNA con entradas en los retrasos j_1, \dots, j_k y con h neuronas en la única capa oculta. Así la figura representa un modelo RNA $(1, 12; 2)$.

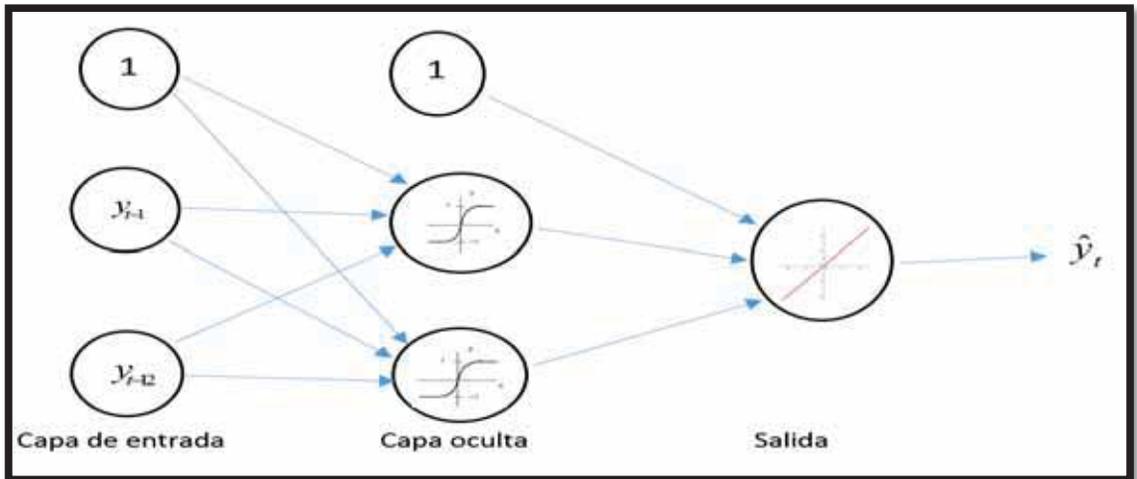


Figura 10: Red Neuronal de alimentación adelante con una capa oculta para la predicción de series de tiempo.

El problema de predicción es formulado como una tarea en la cual el conjunto de unidades de entrada está compuesto por los valores rezagados de la serie de tiempo, y la salida es la predicción del valor siguiente en la secuencia temporal (pronóstico un paso adelante), de tal forma que el comportamiento del sistema se rige por la ecuación:

$$y_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-i})$$

Para representar una serie de tiempo mediante redes perceptrón multicapa, se establece el valor de la serie en el instante en el tiempo t y sus N valores rezagados como entradas a la red en el mismo tiempo t . Se establece, a su vez, el valor de salida de la red como el valor de la serie de tiempo en el instante $t+1$. De este modo, el entrenamiento de la red neuronal minimiza una medida de error entre la salida de la red, para cualquier instante t , y el valor de la serie en el instante $t+1$. Una vez modelada la serie de tiempo mediante la red, para realizar la predicción de los

valores futuros un paso adelante bastará con utilizar las propias salidas de la red como entradas en el tiempo siguiente, de modo que se sigan cumpliendo las ecuaciones iniciales supuestas:

$$y_{t+1} = f(y_t, y_{t-1}, \dots, y_{t-i})$$

$$y_{t+2} = f(y_{t+1}, y_t, \dots, y_{t-i+1})$$

⋮

Algoritmo de Retropropagación o del gradiente descendente: Es la generalización del algoritmo para perceptrón continuo cuando se tiene múltiples capas. Popularmente utilizados para la solución de problemas de clasificación y pronóstico.

El uso del término “retropropagación” aparece después de 1985, sin embargo, la idea básica de retropropagación fue descrita primero por Werbos en su tesis doctoral en 1974. Subsecuentemente, fue redescubierto por Rumelhart, Hinton y Williams en 1986, y popularizado mediante la publicación del libro titulado “Parallel Distributed Processing” por los autores Rumelhart y McClelland. Una generalización similar del algoritmo fue derivada independientemente por (Lecum 1985), y (Haykin, S.1994).

El método general de entrenamiento puede resumirse en los siguientes pasos:

Pasos hacia delante:

- a. Seleccionar un patrón de entrada del conjunto de entrenamiento.
- b. Aplicar esta entrada a la red y calcular la salida.

Pasos hacia atrás:

- a. Calcular el error entre la salida de la red neuronal y la salida deseada para el patrón de entrada usado.
- b. Ajustar los pesos para que el error cometido entre la salida de la red neuronal y la salida deseada sea disminuida.
- c. Repetir los pasos 1 al 4 para todos los patrones de entrenamiento, hasta que el error global sea aceptablemente bajo.

Los valores de los pesos se buscan en forma tal que se reduzca la función del error (diferencia cuadrática entre la salida deseada y la salida de la red neuronal), y este proceso se realiza por el método del gradiente descendente. La idea del método es realizar un cambio en los pesos inversamente proporcionales a la derivada del error respecto al peso para cada patrón.

Durante la fase de asociación o clasificación, la red opera enteramente en forma de cascada directa o hacia delante sin retroalimentación. Sin embargo, “el ajuste de los pesos obtenido por la regla de entrenamiento se realiza desde atrás hacia delante, pasando por las capas ocultas hasta el nivel de entrada” (Colina, E. & Rivas, F. 1998). Como el error se propaga en la red de atrás hacia delante, se le denominó a este procedimiento algoritmo de retropropagación.

Algoritmo de propagación hacia atrás (Backpropagation)

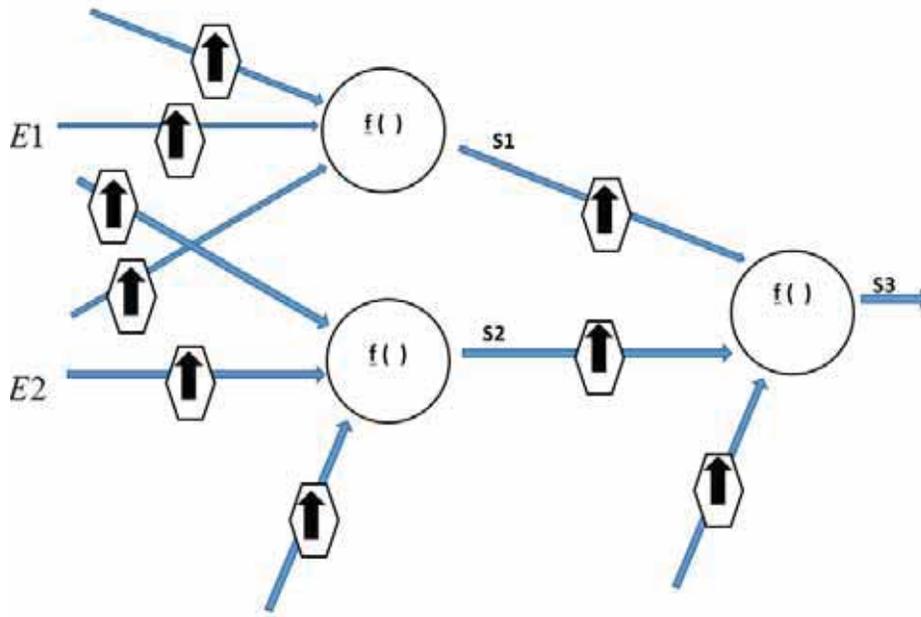


Figura 11: conexión entre tres neuronas

E1 y E2 son las entradas externas, los valores que da el problema. Se observa que E1 entra con un peso en la neurona de arriba y con otro peso en la neurona de abajo. Sucede lo mismo con la entrada E2. Tanto la neurona de arriba como la de abajo tienen sus propias entradas internas.

Lo interesante viene después, porque la salida de la neurona de arriba que es S1 y la salida de la neurona de abajo que es S2 se convierte en entradas para la neurona de la derecha y esas entradas a su vez tienen sus propios pesos. Al final el sistema genera una salida S3 que es la respuesta final de la red neuronal.

¿Qué importante tiene eso? Que si queremos ajustar S3 al resultado que esperamos, entonces retrocedemos a las entradas (S1 y S2) de esa neurona de la derecha ajustando sus pesos respectivos y por supuesto, ese ajuste nos hace

retroceder más aun hasta mirar los pesos de las neuronas de arriba y abajo. Eso se conocerá como el algoritmo de propagación hacia atrás de errores o retro propagación (backpropagation).

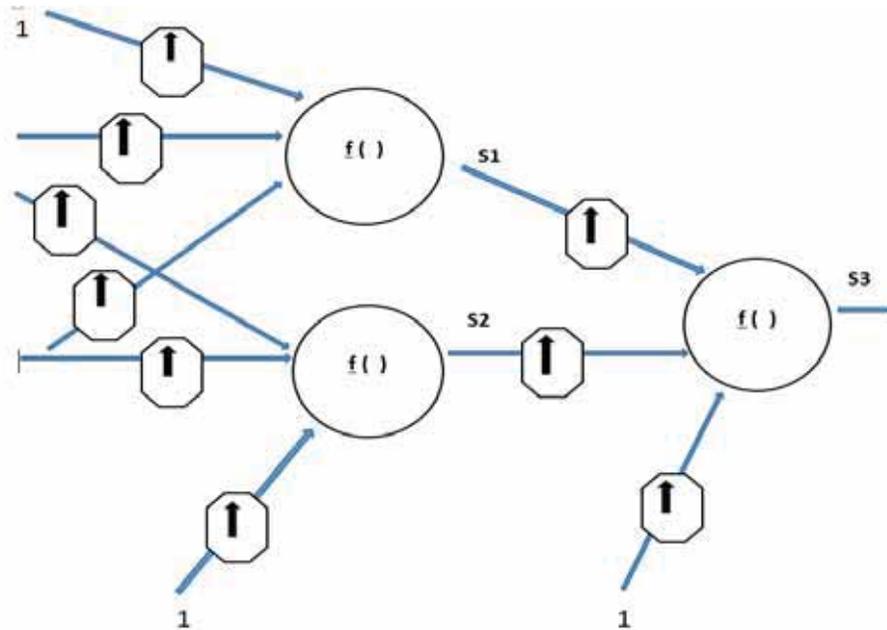


Figura 12: Backpropagation

Luego:

$$S1 = f(E1 \times P1 + E2 \times P3 + 1 \times U1)$$

$$S2 = f(E1 \times P2 + E2 \times P4 + 1 \times U2)$$

$$S3 = f(S1 \times P5 + S2 \times P6 + 1 \times U3)$$

Se concluye entonces que

$$S3 = f(f(E1 \times P1 + E2 \times P3 + 1 \times U1) \times P5 + f(E1 \times P2 + E2 \times P4 + 1 \times U2) \times P6 + 1 \times U3)$$

Error de Entrenamiento

Para medir el error de entrenamiento comúnmente se utiliza la suma cuadrática del error (SSE_e) de entrenamiento, cuya expresión matemática es:

$$SSE_e = (Y_d - Y_n)^2$$

Donde:

Y_d : Y deseada, es la salida del modelo, especificada en el patrón de entrenamiento.

Y_n : Y neuronal, es la salida proporcional por la red, cuando ésta ya ha sido entrenada y se le proporcionan las entradas del patrón de entrenamiento.

Generalización

Una vez que una red neuronal ha sido entrenada mediante un algoritmo supervisado, esto implica que los pesos de interconexión han sido ajustados de forma que se minimice el error. Esta red deberá estar en capacidad de proveer salidas próximas a los valores deseados cuando se le proporcionen nuevos ejemplos, es decir, entradas que no pertenecen al conjunto de entrenamiento, sino que forman parte de los patrones de prueba. A este proceso se le conoce como generalización de un modelo de red neuronal artificial (RNA).

Para que ocurra una buena generalización deben cumplirse las siguientes condiciones:

- Las entradas de la red deben contener suficiente información en relación con el objetivo o salida deseada, de manera que exista una función matemática que relacione correctamente (con el grado de exactitud deseado) las salidas con las entradas.

$$\Gamma : X \rightarrow Y$$

$$\Gamma(x_i) = y_j$$

- La función que se trata de aprender debe ser, en cierto sentido, suave. Es decir, cambios pequeños en las entradas, producen cambios pequeños en las salidas.
- El número de patrones de entrenamiento deberá ser suficientemente grande y representativo del conjunto de los casos que se quiere generalizar.

Error de Generalización

Para medir el error de generalización comúnmente se utiliza la suma cuadrática del error de generalización (SSE_g), cuya expresión matemática es:

$$SSE_g = (Y_d - Y_n)^2$$

Donde:

Y_d : Y deseada, es la salida del modelo, especificada en el patrón de prueba.

Y_n : Y neuronal, es la salida proporcional por la red, cuando ésta ya ha sido entrenada previamente y se le proporcionan las entradas del patrón de prueba.

2.3. Marco Conceptual

- Autocorrelación.** “Correlación que existe entre una variable retrasada uno o más periodos consigo misma” (Hanke, J.E., & Wichern, D.W. 2006).
- Coefficiente de Correlación.** “Mide la magnitud de la relación lineal de dos variables entre sí” (Hanke, J.E., & Wichern, D.W. 2006).

- c. **Componente cíclico.** “Fluctuación parecida a olas alrededor de la tendencia” (Hanke, J.E., & Wichern, D.W. 2006).
- d. **Componente estacional.** “Patrón de cambios que se repite año tras año” (Hanke, J.E., & Wichern, D.W. 2006).
- e. **Correlograma o función de autocorrelación.** “Gráfica de las autocorrelaciones para varios retrasos de una serie de tiempo” (Hanke, J.E., & Wichern, D.W. 2006).
- f. **Corte transversal.** “Observaciones recolectadas en único periodo” (Hanke, J.E., & Wichern, D.W. 2006).
- g. **Residual.** “Diferencia entre el valor real y su valor de pronóstico” (Hanke, J.E., & Wichern, D.W. 2006).
- h. **Serie de tiempo.** “Consiste en datos que se recopilan, registran u observan a lo largo de incrementos sucesivos de tiempo” (Hanke, J.E., & Wichern, D.W. 2006).
- i. **Serie estacionaria.** “Aquella cuyas particulares estadísticas fundamentales, como la media y la varianza, permanecen constante a lo largo del tiempo” (Hanke, J.E., & Wichern, D.W. 2006).
- j. **Tendencia.** “Componente de largo plazo que representa el crecimiento o decremento en las series de tiempo durante un amplio periodo” (Hanke, J.E., & Wichern, D.W. 2006).
- k. **Arquitectura o Modelo.** “La distribución de células en una red neuronal” (Martín Martín, Q. & Rosario de Paz Santana, Y.2007).

- I. **Célula o Neurona.** “Un cálculo lineal o no lineal de elementos que acepta una o más entradas, evalúa una función, y puede dirigir el resultado a una o más células” (Martín Martín, Q. & Rosario de Paz Santana, Y. 2007).
- m. **Pesos.** “Los parámetros de la red que son determinados por entrenamiento iterativo” (Martín Martín, Q. & Rosario de Paz Santana, Y. 2007).

CAPÍTULO III

HIPÓTESIS Y VARIABLES

3.1. Hipótesis

3.1.1. Hipótesis general

El modelo de Redes Neuronales Artificiales pronóstica adecuadamente el número de visitantes extranjeros a Machu Picchu 2002-2018 en comparación a la metodología de Box y Jenkins.

3.1.2. Hipótesis específicas

- El modelo de redes neuronales artificiales es adecuado para el pronóstico del número de visitantes extranjeros a Machu Picchu 2002-2018.
- Los pronósticos del número de visitantes extranjeros a Machu Picchu 2002-2018 son adecuados mediante un modelo de redes neuronales artificiales a diferencia de la metodología de Box y Jenkins.
- Los resultados del número de visitantes extranjeros a Machu Picchu 2002-2018 son similares mediante el modelo de redes neuronales artificiales y la metodología de Box y Jenkins.

3.2. Operacionalización de variables

Tabla 2:

Operacionalización de variables

Tipo de variable	Variable	Indicador(es)	Unidades o Categorías	Escala
Variable dependiente	Número de visitantes extranjeros	Número total mensual de visitantes extranjeros	Número de personas extranjeras	Razón
Variable independiente	Número de visitantes extranjeros retrasados en diferentes periodos de tiempo	Número total mensual de visitantes extranjeros retrasados en diferentes periodos de tiempo	Número de personas extranjeras	Razón

CAPÍTULO IV

METODOLOGÍA

4.1. Tipo y Diseño de Investigación

- Tipo de Investigación: es explicativo, exploratorio y prescriptiva
- Diseño de Investigación: no experimental.

4.2. Población

La población de este estudio comprende la secuencia del número total mensual histórica de visitantes extranjeros al Santuario Histórico de Machu Picchu de Cusco, que están ordenadas y equidistantes en diferentes momentos en el tiempo.

4.3. Muestra no probabilística

El tamaño de la muestra es la serie del número total mensual de visitantes extranjeros al Santuario Histórico de Machu Picchu de Cusco, está compuesta desde el mes de Enero de 2002 al mes de Abril de 2018, por conveniencia del investigador.

4.4. Técnicas de recolección de datos e información

La recopilación de datos para el presente trabajo de investigación, se obtuvo mediante la página web del Ministerio de Comercio Exterior y Turismo (MINCETUR) en la sección estadísticas, en donde el archivo está disponible en formato de Excel, siendo los datos correspondientes al número total mensual de visitantes nacionales y extranjeros al Santuario Histórico de Machu Picchu de Cusco con información de Enero del 2002 a Abril de 2018, en la cual se utilizó los datos del número total mensual de visitantes extranjeros para el análisis correspondiente.

4.5. Métodos de Tratamiento de Datos

Metodología para predicción con Redes Neuronales

A continuación se especifica los pasos a seguir para construir, entrenar y probar una red neuronal para predecir valores futuros de una serie de tiempo, basada únicamente en sus valores pasados.

Escalamiento de los datos

Transformar los datos a valores comprendidos entre 0 y 1, utilizando la siguiente fórmula:

$$Z_t = \frac{y_t}{y_{\max}}$$

Box, Jenking, & Reinsel, (1994) donde:

y_t : son los valores originales de la serie de tiempo

y_{\max} : Valor máximo de la serie de tiempo

Z_t : Serie de tiempo transformada a valores entre 0 y 1

Patrones de entrenamiento y Prueba

Los valores de la serie de tiempo se dividen en dos conjuntos de datos:

- Patrones de entrenamiento: Está formado por el 80% de los datos de la serie. Se seleccionan en forma consecutiva y ordenada. Este conjunto de datos es el que se utilizará para el entrenamiento de la red neuronal.
- Patrones de prueba: Está formado por el 20% de los datos de la serie. Corresponden a los datos restantes, una vez que se han seleccionado los

patrones de entrenamiento. Este conjunto de datos se utiliza para evaluar la capacidad de generalización o predicción de la red.

Topología de las Redes Neuronales Artificiales

- Dirección de la información: Alimentación adelantada
- Tipo de interconexión: Totalmente conectada
- N° de entradas: $p+1$ (una constante de valor 1, denominado sesgo o intercepto)
- N° de capas ocultas: 1
- N° de nodos en la capa oculta: q
- N° de salidas: 1
- Función de activación de los nodos de la capa oculta: logística
- Función de activación de la salida: logística

La función logística siendo una de las más aplicadas en investigaciones, por Khoob (2008) de estructura, perceptron multicapa con algoritmo backpropagation a datos mensuales ; (Polanco C. F., Alvarez G..D. & Moreno V. V. 2015) y también por la forma de su derivada. La elección de otras funciones se realiza según el criterio del investigador, en muchas ocasiones se selecciona por ensayo y error.

Determinación de las entradas (p)

Pueden considerarse varias recomendaciones que ayudarán en la selección de las entradas a la red neuronal artificial (RNA):

- La periodicidad de los datos: Como en esta investigación se utilizan únicamente series de tiempo mensuales, es lógico pensar en considerar 12 o 13 retrasos.

“Podiera ser conveniente construir una primera red con 13 entradas, correspondientes a los 13 retrasos y analizar los pesos, de manera que se seleccionen las entradas asociadas a los pesos de mayor magnitud, como lo sugieren” (Faraway, J. & Chatfield, C.1998). Este análisis de los pesos ayuda a identificar las variables de entrada más importantes.

- Una vez determinado el modelo ARIMA, seleccionar como entradas a los valores correspondientes a los retrasos de Y_t involucrados en este modelo.
- Otra herramienta que podría considerarse son los correlogramas simple y parcial de la serie estacionaria. Las entradas a la red serían los datos correspondientes a los retrasos que resultaran con una correlación significativamente diferente de cero.
- Pruebas por ensayo y error: Otras pruebas pudieran ser de interés, por ejemplo, considerar los datos retrasados 1,4,8,12 periodos, 1,3,6,9,12 periodos, o 1,2,3,4 periodos.

Determinación del número de nodos de la capa oculta (q)

- Una regla valida, en estudios previos ha resultado de utilidad, es asumir que el valor inicial del número de nodos de la capa sea igual al promedio entre el número de entradas y salidas, es decir:

$$\Gamma_n = \frac{x_n + y_k}{2}$$

(Si el valor obtenido es decimal se redondea).

- Pueden realizarse pruebas por ensayo y error, agregando más nodos, y comparando los errores de ajuste y predicción.

Algoritmo de Entrenamiento: Retropropagación

Selección de los pesos iniciales

“La elección de los pesos iniciales puede ser crucial y es recomendable probar con diferentes conjuntos de valores iniciales para tratar de obtener buenos resultados. Los pesos iniciales se generan aleatoriamente 50 veces” (Faraway, J. & Chatfield, C.1998). Se selecciona el modelo que obtenga el menor promedio entre la suma de cuadrados de los errores de ajuste y predicción.

Entrenamiento de la Red Neuronal seleccionada

Para entrenar la red es necesario establecer los siguientes parámetros:

- ✓ El número máximo de ciclos y el error permitido de convergencia se fijará por ensayo y error.
- ✓ Tasa de aprendizaje, incremento de la tasa de aprendizaje y momento pueden ser fijados en 0.05; 1.05 y 0.95, respectivamente. Es conveniente realizar pruebas cambiando estos valores, y evaluando el comportamiento de los errores de entrenamiento y generalización.
- Una vez definida la red neuronal artificial (RNA), con su ecuación se generan los valores de la serie de tiempo ajustada o producida por la red, utilizando los patrones de entrenamiento.
- Se calcula el error de entrenamiento.

Predicción

Usando la ecuación de predicción definida por la RNA se obtiene el valor de predicción $t+1$. Para hacer predicciones más allá del periodo $t+1$, se utiliza ésta como entrada para producir la predicción $t+2$ y así sucesivamente para todo el conjunto de predicción.

Software utilizado

En el presente trabajo de investigación se utilizó el lenguaje de programación R, que es multiplataforma y además de un conjunto de paquetes complementarios del mismo para el análisis, cálculo y gráfico de la información.

CAPÍTULO V

RESULTADOS

5.1. Procesamiento, análisis e interpretación de los datos

5.1.1. Método de Box y Jenkins.

Del procesamiento de los datos se analiza la serie de tiempo número mensual de visitantes extranjeros desde el mes de Enero de 2002 al mes de Diciembre de 2017 para el ajuste de los modelos y los cuatro meses restante (desde Enero de 2018 a Abril de 2018) se utilizaron para ver la precisión de los pronósticos a corto plazo mediante el método de Box y Jenkins.

5.1.1.1. Identificación del modelo.

A través de la gráfica de series de tiempo se observa y analiza lo siguiente:

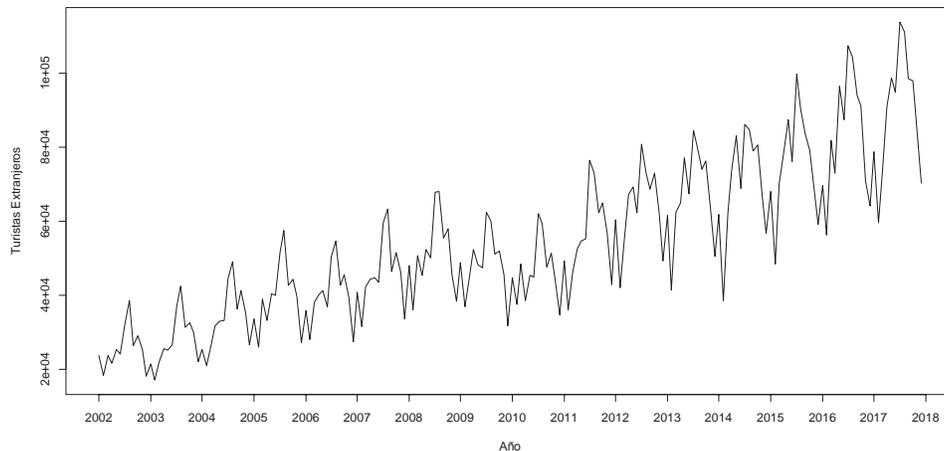


Figura 13: Gráfico de secuencia de la serie número mensual de visitantes extranjeros.

La evolución de la serie de tiempo (ver figura 13) donde se observa que la serie no es estacionaria, se repite año tras año con una tendencia ascendente a lo largo del

tiempo el cual no presenta una media y varianza constante (heterocedasticidad) a lo largo del tiempo, sin embargo tuvo un descenso importante durante los años 2009 y 2010, por lo cual el modelo de componentes multiplicativos es el más adecuado para la serie de tiempo.

$$Y(t) = T(t) \cdot E(t) \cdot C(t) \cdot A(t)$$

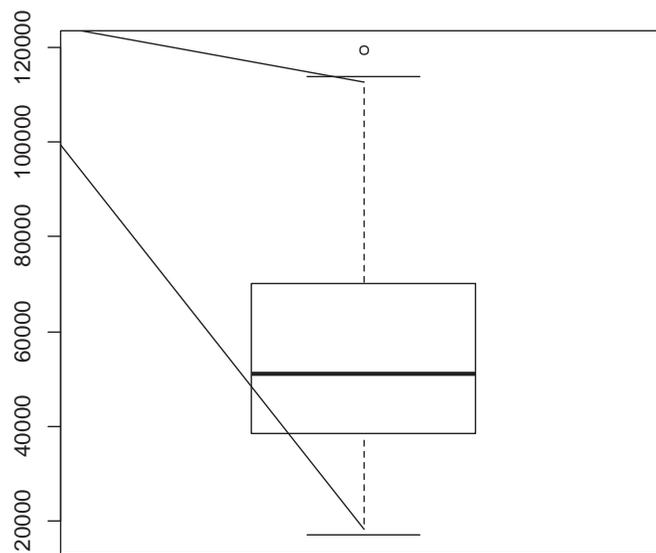


Figura 14: Diagrama de caja de la serie número mensual de visitantes extranjeros

Se observa que los datos de la serie no presenta mucha variabilidad (altura de la caja), se tiene un dato atípico fuera de los límites permitidos; según la prueba de Shapiro Wilk ($W = 0.9653$, $p \text{ valor} = 8.468e-05$) nos reporta que los datos no tienen una aproximación a una distribución normal es decir sobre el sesgo hacia número de visitantes menores.

(Prueba de normalidad - prueba de Shapiro-Wilks: se obtiene el valor del estadístico y el valor P de probabilidad del contraste. Se rechaza H_0 si el valor P de

probabilidad es menor que el nivel de significación elegido para ejecutar la prueba de contraste estadístico (www.uv.es/friasnav/).

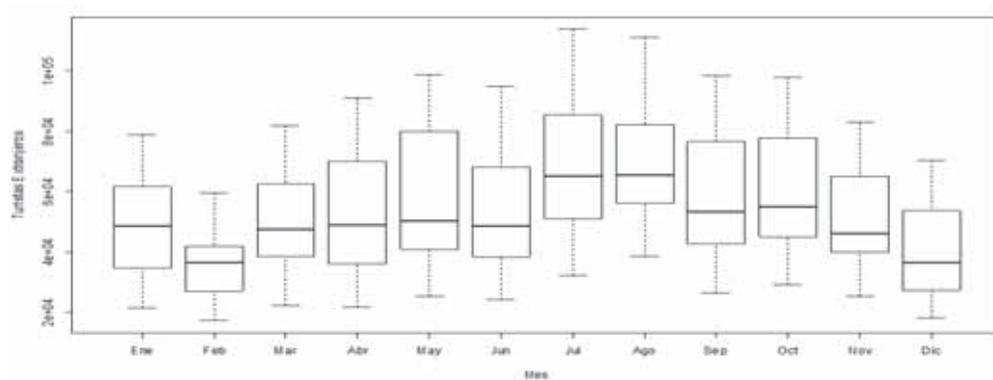


Figura 15: Diagrama de cajas por mes de los años de la serie número mensual de visitantes extranjeros.

Fuente: Elaboración propia

Se observa un análisis mediante diagramas de cajas por meses, para ver el comportamiento estacional (ver figura 15), en donde se puede ver que se tiene variabilidad en la mayoría de las cajas, también la mediana en su mayoría de las cajas no está centrada, esto en vista que la serie presenta una tendencia creciente a lo largo del tiempo; por otra parte en los meses de julio y agosto se tuvieron las mayores demandas en visitas de turistas extranjeros en promedio, en febrero y diciembre se registraron las menores demandas de turistas extranjeros.

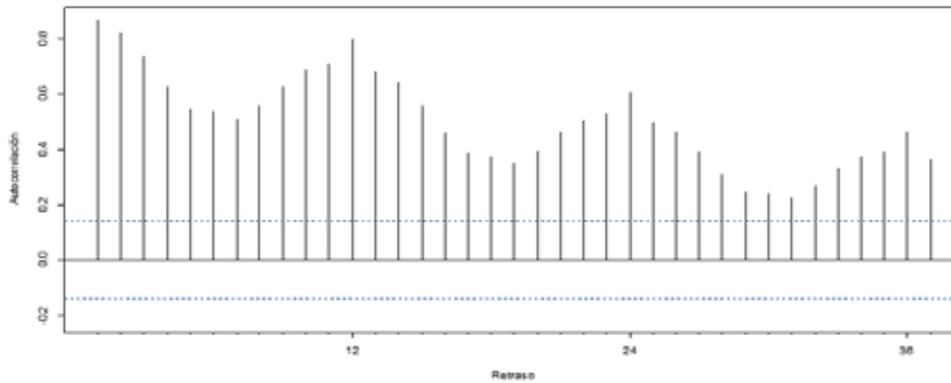


Figura 16: Función de Autocorrelación de la serie número mensual de visitantes extranjeros.

Se observa las autocorrelaciones (ver figura 16) de la muestra, donde cada una eran persistentemente grandes y se desvanecían hacia cero lentamente, también se observa que las autocorrelaciones en los retrasos 12, 24 y 36 son grandes y no se desvanecían con rapidez, se decidió diferenciar la serie en relación con el retraso estacional para convertir la serie no estacionaria en una que si lo fuera.

Tabla 3:

Comparación de modelos tentativos para la serie número mensual de visitantes extranjeros

Modelo	Medidas de ajuste			Precisión de pronóstico	
	RMSE	MAPE	AIC	RMSE	MAPE
ARIMA (1,0,1) (0,1,1) ₁₂	3688.885	5.116751	3492.41	9064.077	7.656144
ARIMA (2,0,1) (0,1,1) ₁₂	3688.065	5.101802	3494.29	9086.574	7.658852
ARIMA (1,0,2) (0,1,1) ₁₂	3688.061	5.100868	3494.28	9089.471	7.661061
ARIMA (2,0,2) (0,1,1) ₁₂	3686.564	5.102333	3496.19	9108.812	7.693333
ARIMA (1,0,1) (2,1,1) ₁₂	3527.566	4.941737	3483.22	8952.232	7.625743

MAPE: Error porcentual absoluto medio; RMSE: Raíz cuadrada media del error; AIC: Criterio de Akaike.

La tabla N° 3, nos muestra cinco modelos candidatos a elegir, con menores AIC y indicadores estadísticos (RMSE, MAPE).

La función "Arima" de la librería "forecast" es el que se utilizó para la estimación de los modelos (ver anexo 1).

Construcción de los modelos de la tabla N° 3, son:

ARIMA (1,0,1) (2,1,1)¹² con AIC = 3483.22

$$(1-\phi_1 L) \left[1 - (\Phi_1 L^{12} + \Phi_2 L^{24}) \right] (1-L^{12}) y_t = (1+\theta_1 L)(1+\Theta_1 L^{12}) \varepsilon_t$$

$$(1-\phi_1 L)(1-\Phi_1 L^{12} - \Phi_2 L^{24})(1-L^{12}) y_t = (1+\theta_1 L)(1+\Theta_1 L^{12}) \varepsilon_t$$

$$(1-\Phi_1 L^{12} - \Phi_2 L^{24} - \phi_1 L + \Phi_1 \phi_1 L^{13} + \Phi_2 \phi_1 L^{25} - L^{12} + \Phi_1 L^{24} + \Phi_2 L^{36} + \phi_1 L^{13} - \Phi_1 \phi_1 L^{25} - \Phi_2 \phi_1 L^{37}) y_t = (1+\Phi_1 L^{12} + \theta_1 L + \Theta_1 \theta_1 L^{13}) \varepsilon_t$$

$$(1-\Phi_1 L^{12} - \Phi_2 L^{24} - \phi_1 L + \Phi_1 \phi_1 L^{13} + \Phi_2 \phi_1 L^{25} - L^{12} + \Phi_1 L^{24} + \Phi_2 L^{36} + \phi_1 L^{13} - \Phi_1 \phi_1 L^{25} - \Phi_2 \phi_1 L^{37}) y_t = (1+\Phi_1 L^{12} + \theta_1 L + \Theta_1 \theta_1 L^{13}) \varepsilon_t$$

$$y_t - \Phi_1 L^{12} y_t - \Phi_2 L^{24} y_t - \phi_1 L y_t + \Phi_1 \phi_1 L^{13} y_t + \Phi_2 \phi_1 L^{25} y_t - L^{12} y_t + \Phi_1 L^{24} y_t + \Phi_2 L^{36} y_t + \phi_1 L^{13} y_t - \Phi_1 \phi_1 L^{25} y_t - \Phi_2 \phi_1 L^{37} y_t = \varepsilon_t + \Phi_1 L^{12} \varepsilon_t + \theta_1 L \varepsilon_t + \Theta_1 \theta_1 L^{13} \varepsilon_t$$

$$y_t - \Phi_1 y_{t-12} - \Phi_2 y_{t-24} - \phi_1 y_{t-1} + \Phi_1 \phi_1 y_{t-13} + \Phi_2 \phi_1 y_{t-25} - y_{t-12} + \Phi_1 y_{t-24} + \Phi_2 y_{t-36} + \phi_1 y_{t-13} - \Phi_1 \phi_1 y_{t-25} - \Phi_2 \phi_1 y_{t-37} = \varepsilon_t + \Phi_1 \varepsilon_{t-12} + \theta_1 \varepsilon_{t-1} + \Theta_1 \theta_1 \varepsilon_{t-13}$$

$$y_t - \phi_1 (y_{t-1} - \Phi_1 y_{t-13} - \Phi_2 y_{t-25} - y_{t-13} + \Phi_1 y_{t-25} + \Phi_2 y_{t-37}) - \Phi_1 y_{t-12} - \Phi_2 y_{t-24} - y_{t-12} + \Phi_1 y_{t-24} + \Phi_2 y_{t-36} = \varepsilon_t + \theta_1 (\varepsilon_{t-1} + \Theta_1 \varepsilon_{t-13}) + \Theta_1 \varepsilon_{t-12}$$

$$y_t = \phi_1 (y_{t-1} - \Phi_1 y_{t-13} - \Phi_2 y_{t-25} - y_{t-13} + \Phi_1 y_{t-25} + \Phi_2 y_{t-37}) + \Phi_1 y_{t-12} + \Phi_2 y_{t-24} + y_{t-12} - \Phi_1 y_{t-24} - \Phi_2 y_{t-36} + \varepsilon_t + \theta_1 (\varepsilon_{t-1} + \Theta_1 \varepsilon_{t-13}) + \Theta_1 \varepsilon_{t-12}$$

Siendo este modelo con menor AIC de la metodología Box-Yenkins.

ARIMA (2,0,2) (0,1,1)₁₂ con AIC = 3496.19

$$\begin{aligned}
& \left[1 - (\phi_1 L + \phi_2 L^2) \right] \left[(1 - L^{12}) \right] y_t = (1 + \Theta L^2)(1 + \Theta L + \Theta L) \varepsilon_t \\
& (1 - \phi_1 L - \phi_2 L^2)(1 - L^{12}) y_t = (1 + \theta_1 L + \theta_2 L^2 + \Theta_1 L^{12} + \Theta_1 \theta_1 L^{13} + \Theta_1 \theta_2 L^{14}) \varepsilon_t \\
& (1 - \phi_1 L - \phi_2 L^2 - L^{12} + \phi_1 L^{13} + \phi_1 L^{14}) y_t = \varepsilon_t + \theta_1 L \varepsilon_t + \theta_2 L^2 \varepsilon_t + \Theta_1 L^{12} \varepsilon_t + \Theta_1 \theta_1 L^{13} \varepsilon_t + \Theta_1 \theta_2 L^{14} \varepsilon_t \\
& y_t - \phi_1 L y_t - \phi_2 L^2 y_t - L^{12} y_t + \phi_1 L^{13} y_t + \phi_1 L^{14} y_t = \varepsilon_t + \theta_1 L \varepsilon_t + \theta_2 L^2 \varepsilon_t + \Theta_1 L^{12} \varepsilon_t + \Theta_1 \theta_1 L^{13} \varepsilon_t + \Theta_1 \theta_2 L^{14} \varepsilon_t \\
& y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - y_{t-12} + \phi_1 y_{t-13} + \phi_2 y_{t-14} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \Theta_1 \varepsilon_{t-12} + \Theta_1 \theta_1 \varepsilon_{t-13} + \Theta_1 \theta_2 \varepsilon_{t-14} \\
& y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \Theta_1 \varepsilon_{t-12} + \Theta_1 \theta_1 \varepsilon_{t-13} + \Theta_1 \theta_2 \varepsilon_{t-14} + \phi_1 y_{t-1} + \phi_2 y_{t-2} + y_{t-12} - \phi_1 y_{t-13} - \phi_2 y_{t-14} \\
& y_t = \varepsilon_t + \theta_1 (\varepsilon_{t-1} + \Theta_1 \varepsilon_{t-13}) + \theta_2 (\varepsilon_{t-2} + \Theta_1 \varepsilon_{t-14}) + \Theta_1 \varepsilon_{t-12} + \phi_1 (y_{t-1} - y_{t-13}) + \phi_2 (y_{t-2} - y_{t-14}) + y_{t-12}
\end{aligned}$$

Siendo este modelo con mayor AIC, de la metodología Box-Yenkins.

ARIMA (1,0,1)(0,1,1)₁₂ con AIC = 3492.41

$$\begin{aligned}
 [1 - \phi_1 L](1)(1)(1 - L^{12})y_t &= (1 + \theta_1 L)(1 + \Theta_1 L^{12})\varepsilon_t \\
 (1 - \phi_1 L)(1 - L^{12})y_t &= (1 + \theta_1 L)(1 + \Theta_1 L^{12})\varepsilon_t \\
 (1 - L^{12} - \phi_1 L + \phi_1 L^{13})y_t &= (1 + \Theta_1 L^{12} + \theta_1 L + \theta_1 \Theta_1 L^{13})\varepsilon_t \\
 y_t - y_{t-12} - \phi_1 y_{t-1} + \phi_1 y_{t-13} &= \varepsilon_t + \Theta_1 \varepsilon_{t-12} + \theta_1 \varepsilon_{t-1} + \theta_1 \Theta_1 \varepsilon_{t-13} \\
 y_t &= \phi_1 (y_{t-1} - y_{t-13}) + y_{t-12} + \theta_1 (\varepsilon_{t-1} + \Theta_1 \varepsilon_{t-13}) + \Theta_1 \varepsilon_{t-12}
 \end{aligned}$$

Siendo este modelo utilizado el segundo con menor AIC de los cinco modelos candidatos, con índices estadísticos menores y con pocos parámetros para su construcción como se observa.

5.1.1.2. Estimación del modelo.

Se realizaron numerosas pruebas, siendo el modelo ARIMA (1,0,1) (0,1,1)₁₂ dado que su construcción es menos laborioso (tiene menos parámetros) “principio de parsimonia”, con similar ajuste y una precisión en el pronóstico idéntica a los otros modelos, los coeficientes del modelo con constante se muestran en la tabla siguiente:

Tabla 4:
Coeficientes estimados para el modelo seleccionado

Coeficiente	Estimado	Error estándar
ar1	0.9199	0.0399
ma1	-0.5810	0.0793
sma1	-0.4617	0.0663
Constante	345.9181	66.7162

$$W_t = Y_t - Y_{t-12}$$

$$W_t = 345.9181 + 0.9199W_{t-1} + \varepsilon_t - 0.5810\varepsilon_{t-1} - 0.4617\varepsilon_{t-12}$$

$$\hat{Y}_t = 345.9181 + Y_{t-12} + 0.9199(Y_{t-1} - Y_{t-13}) - 0.5810\varepsilon_{t-1} - 0.4617\varepsilon_{t-12}$$

La ecuación de pronóstico es:

$$\hat{Y}_t = 345.9181 + 0.9199Y_{t-1} + Y_{t-12} - 0.9199Y_{t-13} - 0.5810\varepsilon_{t-1} - 0.4617\varepsilon_{t-12}$$

5.1.1.3. Evaluación del modelo

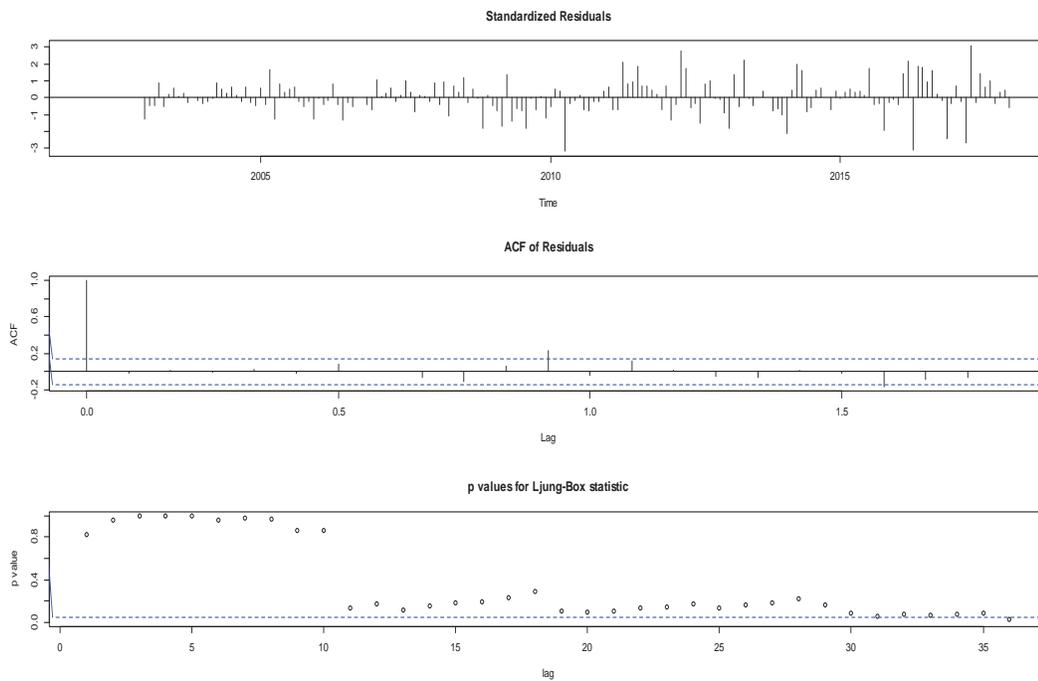


Figura 17: Diagnóstico de los residuos del modelo ajustado

Fuente: Elaboración propia

En la figura 17, presenta los residuos y la autocorrelación de los residuos y la estadística Ljung-Box, conforme se juzgan por los valores individuales, son coherentes con los errores aleatorios, sugiere que el modelo es adecuado, puesto que en su mayoría de estos puntos tienden a cero viéndose a partir del punto 11.

5.1.1.4. Realización de pronósticos con el modelo.

Los pronósticos no fueron completamente coherentes con la conducta de la serie, no captura perfectamente la esencia del crecimiento del número mensual de visitantes extranjeros (ver Tabla 3).

Tabla 5:

Valores reales y pronosticados del número mensual de visitantes extranjeros

Año	Mes	Real	Pronóstico	Li 95	Ls 95
2018	enero	84898	84405	76853.28	91956.43
2018	febrero	63051	66214	58240.37	74187.31
2018	marzo	93276	84123	75809.55	92437.07
2018	abril	102112	93049	84457.53	101639.88
2018	mayo	119363	104774	95955.02	113593.21
2018	junio	108128	97933	88925.08	106939.98

Nota. L.i. 95 = Límite inferior al 95%; L.s. 95 = Límite superior al 95%.

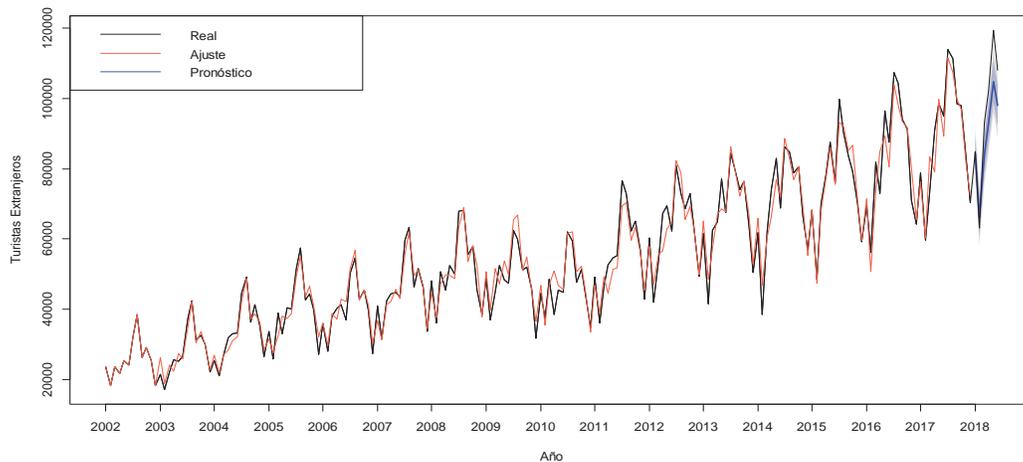


Figura 18: Valores ajustados y pronosticados con límites de confianza al 95% por el modelo $ARIMA(1,0,1)(0,1,1)_{12}$ para serie número de visitantes extranjeros.

Fuente: Elaboración propia

En la figura se observa los valores reales, ajustados y pronósticos para el primer semestre del año 2018.

5.1.2. Método de redes neuronales.

Se sigue los procedimientos especificados en el capítulo IV, para construir un modelo de red neuronal tipo Perceptrón Multicapa (backpropagation), “consiste en su capacidad de organizar una representación interna del conocimiento en las capas ocultas de neuronas, a fin de aprender la relación que existe entre un conjunto de entradas y salidas. Luego aplica esa misma relación a nuevos vectores de entrada con ruido o incompletos, dando una salida activa si la nueva entrada es parecida a las presentadas durante el aprendizaje. Esta característica importante es la capacidad de generalización, entendida como la facilidad de dar salidas satisfactorias a entradas que el sistema no ha visto nunca en su fase de entrenamiento”.

Se utilizó el lenguaje de programación R versión 3.5.1, para desarrollar todos los procedimientos computacionales, entrenamiento y validación de las redes neuronales artificiales.

5.1.2.1. Escalamiento de los datos

Todos los datos fueron transformados dentro del intervalo comprendido entre 0 y 1, mediante procedimientos computacionales.

5.1.2.2. Patrones de entrenamiento y prueba.

Para el entrenamiento se utilizó la información disponible desde el mes de Enero de 2002 al mes de Diciembre de 2017 y para la validación se utilizó los cuatro meses restantes (Enero - Abril del 2018).

Es decir se dividió en dos submuestras, la muestra de entrenamiento y la muestra de prueba. La primera muestra permite entrenar la red; mientras que la muestra de prueba permite corregir y comprobar el entrenamiento.

5.1.2.3. Topología de la red neuronal artificial.

Todas las redes neuronales artificiales son totalmente conectadas, con una capa oculta y una función de activación logística, una neurona (nodo) adicional tanto en la capa de entrada como en la capa oculta de valor constante 1, y solo una neurona (nodo) en la salida más la función de activación logística en vista que transforma valores grandes a valores simples entre 0 y 1.

5.1.2.4. Determinación de las entradas.

Se consideró con variables potenciales significativas en el modelo las primeras 13 entradas (retrasos) de la serie, de tal manera que se considere un ciclo completo del periodo estacional.

Se contó con la ayuda de las funciones de autocorrelación simple y parcial (ver figura 16) para la optimización de las entradas (retrasos), en donde los retrasos 1, 2, 3, 4, 12 y 24 tuvieron una autocorrelación importante diferente de cero.

Las entradas (retrasos), por ejemplo 1, 2 y 12 retrasos, 1, 2, 12 y 24 retrasos y 1, 2, 3, 4, 5, 6, 12 y 24 retrasos fueron considerados por ensayo y error.

5.1.2.5. Determinación del número de neuronas en la capa oculta.

Se consideró de dos hasta ocho neuronas, se realizaron muchas pruebas incluyendo y disminuyendo el número de neuronas (nodos), pero no hubo diferencias significativas.

5.1.2.6. Algoritmo de entrenamiento.

Todas las redes neuronales artificiales se entrenaron con el algoritmo propagación hacia atrás.

5.1.2.7. Selección de los pesos iniciales.

Se generaron 1000 redes neuronales para ajuste, cada uno con pesos iniciales aleatorios. Estos se promedian luego al calcular los pronósticos.

5.1.2.8. Ajuste o entrenamiento de la red neuronal artificial.

Número máximo de iteraciones: 300

Criterio de convergencia: 0.0001

5.1.2.9. Pronostico.

Usando la ecuación de predicción definida por la Red Neuronal Artificial se obtiene el valor de predicción $t+1$. Para hacer predicciones más allá del periodo $t+1$, se utiliza ésta como entrada para producir la predicción $t+2$ y así sucesivamente para todo el conjunto de predicción.

5.1.3. Modelos de redes neuronales para pronóstico.

Los resultados obtenidos referentes al entrenamiento y validación de las diferentes configuraciones de modelos de redes neuronales artificiales se muestran (tabla 6),

para la simulación del modelo de redes neuronales se utilizó la función “nnetar” de la librería “forecast”, el código se puede verificar (anexo 2).

Tabla 6:

Comparación de varios modelos de redes neuronales artificiales y el modelo de Box-Jenkins ajustado a la serie.

Modelo	Retrasos	h	Medidas de ajuste		Precisión de pronóstico	
			RMSE	MAPE	RMSE	MAPE
1	1-13	7	657.816	1.078998	10138.823	9.215359
2	1-13, 24	8	288.608	0.450561	11468.254	10.326260
3	1, 2, 12	2	4682.611	7.161341	10239.092	8.518699
4	1, 2, 12, 24	2	4592.188	6.914733	7944.512	7.004750
5	1, 2, 12, 24	4	4005.589	6.122630	7445.887	6.400256
6	1-4, 12	3	4044.233	6.063447	10456.515	8.662602
7	1-4, 12, 24	4	3467.285	5.058819	5126.577	4.409447
8	1-4, 12, 24	8	1877.107	2.774582	6736.965	4.605459

Nota. h = número de neuronas (nodos) en la capa oculta; las medidas de error a tomar en consideración son los que están en negrita; MAPE: Error porcentual absoluto medio; RMSE: Raíz cuadrada media del error

En la tabla N° 6 se aprecia diferentes medidas de error (RMSE y MAPE) sobre el desempeño para los ocho modelos de redes neuronales para el ajuste y la precisión de pronóstico. Los modelos 1 y 2 tuvieron los mejores ajustes, pero tuvo una capacidad predictiva menor a comparación del modelo obtenido por el método de Box-Jenkins, esto debido a un sobreajustamiento por el exceso de parámetros en el proceso de entrenamiento, los modelos 3, 4, 5 y 6 tuvieron similares medidas de error en el ajuste, de los cuales los modelos 4 y 5 fueron ligeramente mejores en los pronósticos que el modelo seleccionado por el método de Box-Jenkins, se observó que los modelos 7 y 8 obtuvieron los mejores resultados en la precisión de

pronósticos, obteniendo un error porcentual absoluto medio (MAPE) de 4.41% y 4.61% respectivamente. Siendo el modelo 7 el seleccionado para efectuar los pronósticos de la serie número mensual de visitantes extranjeros. El ajuste y los pronósticos por la red neuronal artificial seleccionada se muestran en la siguiente figura:

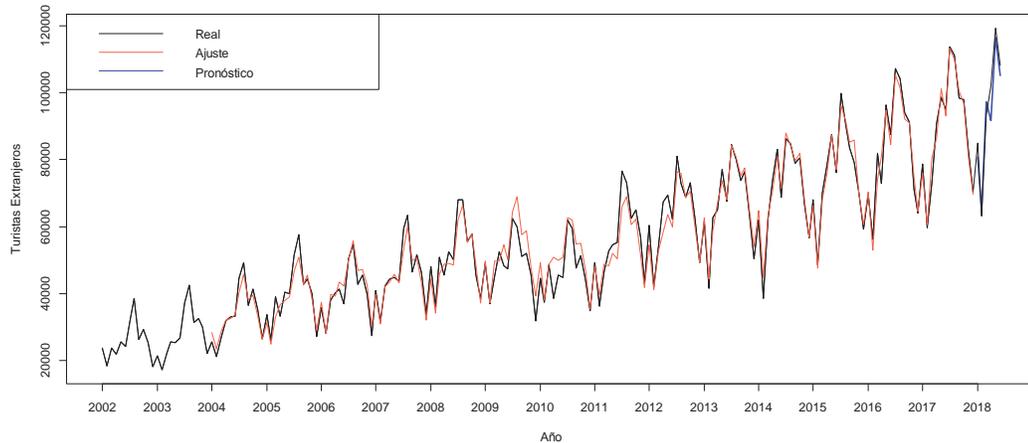


Figura 19: Valores ajustados y pronosticados por la red neuronal artificial modelo 7 para serie número de visitantes extranjeros.

Fuente: Elaboración propia

Tabla 7:

Modelo elegido de la Metodología Box-Yenkins (ARIMA) y Redes Neuronales Artificiales (RNA)

Metodología	Modelo	Medidas de ajuste			Precisión de pronóstico	
		RMSE	MAPE	AIC	RMSE	MAPE
Box-Yenkins	ARIMA (1,0,1)(0,1,1) ₁₂	3688.885	5.116751	3492.41	9064.077	7.656144
RNA	7 1-4, 12, 24	4 3467.285	5.058819		5126.577	4.409447

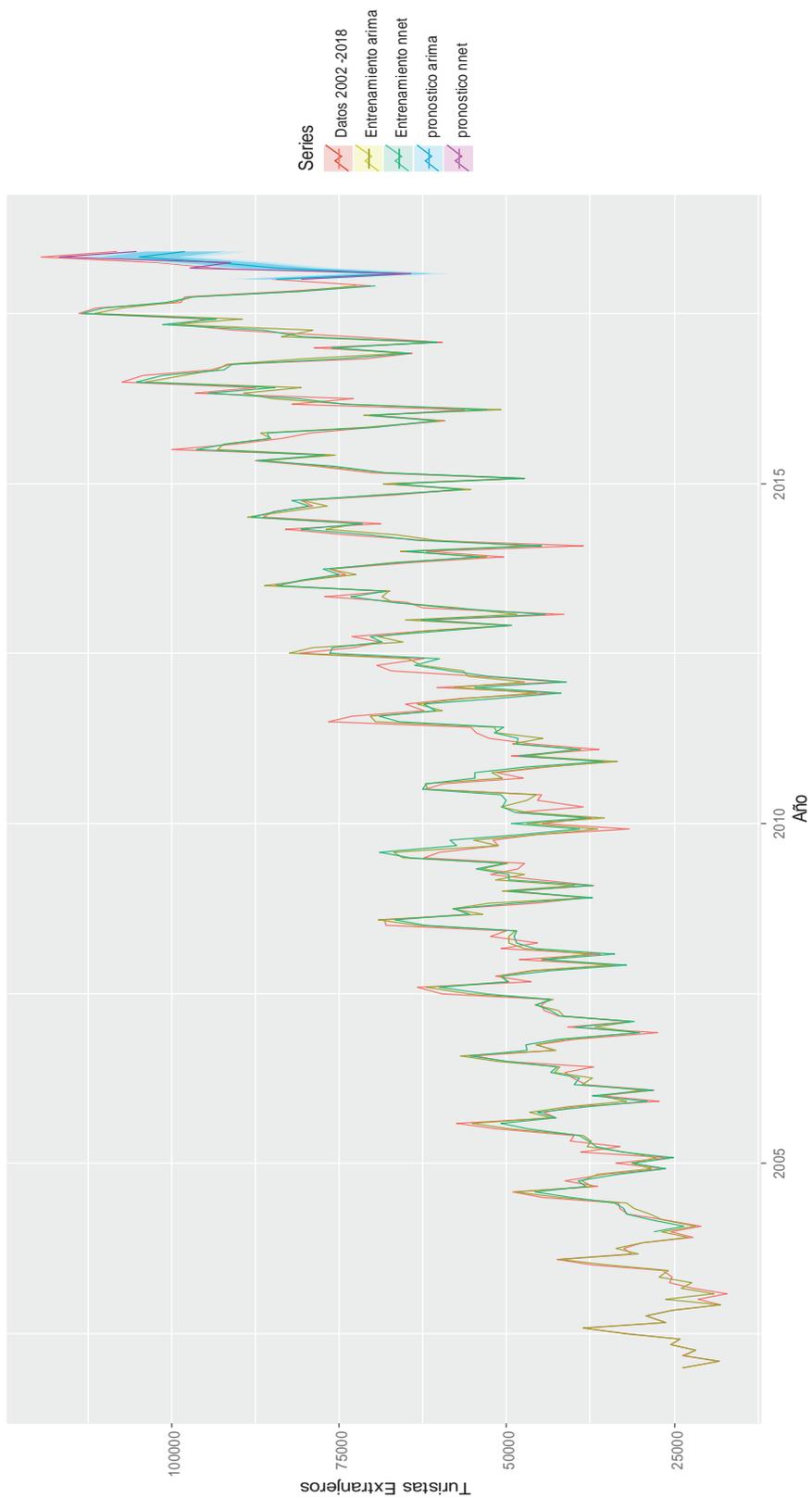


Figura 20: serie con data real, Modelo 1 ARIMA y Modelo 7 de redes neuronales artificiales, pronósticos con ambos modelos.

CONCLUSIONES

PRIMERO: El modelo de red neuronal tipo Perceptron Multicapa es el utilizado contando con seis entradas o retrasos de la serie (Y_{t-1} , Y_{t-2} , Y_{t-3} , Y_{t-4} , Y_{t-12} y Y_{t-24}), los cuales fueron seleccionados por ensayo y error. La función de autocorrelación simple y parcial de la metodología Box y Jenkins sirvió en identificar un óptimo número de entradas, se consideró cuatro neuronas (nodos) en la capa oculta que fueron elegidos por ensayo y error, una neurona (nodo) en la capa de salida (Y_t).

SEGUNDO: Los pronósticos para el primer semestre del 2018 realizados por el modelo de red neuronal artificial capturan adecuadamente la esencia del patrón estacional y el crecimiento del número mensual de visitantes extranjeros, frente a los pronósticos por el modelo ARIMA (1,0,1)(0,1,1)₁₂ para las mismas fechas no resultaron ser tan coherentes con la conducta de la serie; pronósticos logrados sin tomar en cuenta las posibles alteraciones atribuidas a factores externos.

TERCERO: En la serie número mensual de visitantes extranjeros al santuario histórico de Machu Picchu, el modelo 7 de red neuronal artificial consiguió un ajuste menor al modelo ARIMA (1,0,1)(0,1,1)₁₂; a la vez el modelo 7 de (RNA) se eligió para la precisión del pronóstico con índices estadísticos menores en comparación de la metodología de Box-Jenkins (modelo ARIMA).

RECOMENDACIONES

- a. Experimentar con otras series de tiempo como demanda de energía eléctrica, producto bruto interno, producción industrial, etc. Para verificar las ventajas y desventajas del método de redes neuronales artificiales frente a la metodología de Box-Jenkins.
- b. Utilizar el algoritmo óptimo en el entrenamiento de la red neuronal artificial, el cual consigue reducir el tiempo de convergencia.
- c. Utilizar la información de la metodología de Box-Jenkins para identificar el número óptimo de entradas para la red neuronal artificial.
- d. Utilizar el software R-project, en específico la librería forecast y las funciones: nnetar, para la simulación de las redes neuronales artificiales; Arima, para los modelos autorregresivos integrados de media móvil.

BIBLIOGRAFÍA

- Anderson O, D. (1985). *Times Series Analisis and Forecasting*. New York: 1ra Edición.
- Bowerman B, L., & O'connell R, T. (1993). *Pronósticos, Series de Tiempo y regresión*. México: Cengage Learnig.
- Box G, E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*, 211-246.
- Box, Jenkins, & Reinsel. (1994). *Time Series Analysis*. Madrid, NJ: Prentice Hall.
- Camilo S, J. (2006). predicción de series temporales con redes neuronales: una aplicación a la inflación Colombiana. *Revista Colombiana de estadística*, 77.
- Cárdenas G, J. A. (2015). *Pronósticos y comparación de una serie de tiempo con cambios estructurales mediante la red neuronal artificial de Retropropagación Resiliente y modelos no lineales*. Lima - Perú: Universidad Nacional Mayor de San Marcos.
- Colina, E., & Rivas, F. (1998). *Introducción a la inteligencia artificia*. Mérida - Venezuela: Universidad de los Andes.
- Castillo, J. (2015). Concepto de la red neuronal con especial atención al Perceptrón Multicapa como elemento clasificador de patrones. [en línea]. 2017 Editorial, Periódico El Comercio.
- Esteba M, N. (2011). *Técnicas de redes neuronales artificiales para predecir el ingreso de extranjeros al Perú por el Aeropuerto Internacional Jorge Chávez*. Puno: Universidad Nacional del Altiplano.

- Faraway, J. & Chatfield, C. (1998) *Time series forecasting with neural networks: A comparative study using the airline data*. Appl. Statist., 47, 231-250. [3.4, 3.5, 6.4, 8.3]
- Gómez L, W. (2010). *Pronóstico de una serie temporal de demanda máxima de energía eléctrica usando redes neuronales*. Oaxaca: Universidad Tecnológica de la Mixteca.
- Gujarati, D. N. (2004). *Econometría*. México: McGraw-Hill.
- Hagan M, T., Dermuth H, B., & Beale, M. (1996). *Neural Network Desing*. China: Thomson.
- Hanke J, E. (1996). *Pronósticos en los Negocios*. México.
- Hanke, J. E., & Wichern, D. W. (2006). *Pronosticos en los Negocios*. México: Pearson.
- Haykin, S. (1994). *Neural networks: a comprehensive foundation*,. EE.UU: Macmillan College Publishing Company.
- Hyndman , J. R., & Khandakar , Y. (2008). Automatic time series forecasting. *Journal of Statistical Software*, 26.
- Jiménez , U. (1985). *Análisis de Series Temporales – modelos ARIMA*. España: Paraninfo A.A.
- Joekes, S., Barbosa, S., & Robledo, W. (2005). *Modelado y pronóstico de una serie de tiempo contaminada empleando redes neuronales y procedimientos estadísticos tradicionales*.

- Kendall M, G., & Buckland W, R. (1985). *Diccionario de Estadística*. Madrid: Piramide.
- Lecun, Y. (1985) Une producere d'apprentissage pour reseau a seuil assymetrique. *Cognitiva*, Vol.85, pp. 599-604.
- Mamani B, E. (2014). *Redes neuronales artificiales para el pronóstico del número de visitantes extranjeros a la reserva nacional del Titicaca*. Puno: Universidad Nacional del Altiplano.
- Martín Martín Q., & del Rosario Paz Santana, Y (2007). *Aplicación de las redes neuronales artificiales a la regresión*. Madrid: La Muralla, S.A.
- Morales C, A., & Cruz T, I. (2008). *Pronostico mediante redes neuronales artificiales y Modelos ARIMA*. México: Universidad Nacional Autónoma de México (UNAM) .
- Morettin, P. (2002). *Series Temporales en Finanzas*. Sau Paulo - Brazil: Instituto de Matemática y Estadística.
- Palit, A.K., Popovic (2005). *Computational Intelligence in Time series forecasting*. E.E.U.U. Springer.
- Pindick R, S., & Rubinfeld D, L. (2003). *Econometría. Modelos y Pronósticos*. Madrid: McGraw-Hill.
- Uriel, E. (1985) Analisis de series temporales. Modelos ARIMA. Madrid Edita:Paraninfo.

Valencia T. A., & Quispe H. R. (2009). *Modelos no Lineales de Redes Neuronales Artificiales Basadas en la Predicción de Series de Tiempo*. Cusco: Universidad Nacional de San Antonio Abad del Cusco.

- ANEXOS -

- ANEXO 1 -

```
library(forecast) # cargar librería
load("Xt.RData")
Xt # mostrar datos
Yt <- window(Xt, end = c(2017, 12)) # Entrenamiento
lambda <- BoxCox.Lambda (Yt)
fit <- auto.arima (Yt)
fit <- auto.arima (Yt, stepwise = FALSE, approximation = FALSE,
                  lambda = lambda)
tsdiag (fit, 36)
fit <- nnetar (Yt)
accuracy (forecast (fit, h = 6), window(Xt, start = c(2018, 1),
end = c(2018, 6)))
#####
### 1. IDENTIFICACIÓN DEL MODELO
plot(Yt, xlab = "Año", ylab = "Turistas Extranjeros",
     xaxt = "n")
axis (1, seq(2002, 2018, by = 1))
Mes <- c ("Ene", "Feb", "Mar", "Abr", "May", "Jun", "Jul",
         "Ago", "Sep", "Oct", "Nov", "Dic")
Boxplot (Yt ~ cycle(Yt), names = Mes,
         xlab = "Mes", ylab = "Turistas Extranjeros")
Acf (Yt, main = "", lag.max = 37, xlab = "Retraso",
     ylab = "Autocorrelación")
plot (diff (Yt, lag = 12, differences = 1), xlab = "Año",
     ylab = "Turistas Extranjeros diferenciada 12", xaxt = "n")
axis (1, seq(2003, 2018, by = 1))
par(mfrow = c(2, 1), mai = c(0.7, 0.7, 0.6, 0.3), mgp = c(2, 1, 0),
     font.main = 1)
Acf (diff (Yt, lag = 12, differences = 1), main = "a.", lag.max = 37,
```

```

xlab = "Retraso",
ylab = "Autocorrelación")
Pacf (diff (Yt, lag = 12, differences = 1), main = "b.", lag.max = 37,

xlab = "Retraso",

ylab = "Autocorrelación parcial")

```

2. ESTIMACIÓN DE MODELOS

```

m1 <- Arima (Yt, order = c (1, 0, 1), seasonal = c (0, 1, 1),
include.drift = TRUE)
accuracy (forecast (m1, h = 6), window(Xt, start = c(2018, 1),
end = c(2018, 6)))
m2 <- Arima (Yt, order = c(2, 0, 1), seasonal = c(0, 1, 1),
include.drift = TRUE)
accuracy (forecast (m2, h = 6), window (Xt, start = c(2018, 1),
end = c(2018, 6)))
m3 <- Arima(Yt, order = c(1, 0, 2), seasonal = c(0, 1, 1),
include.drift = TRUE)
accuracy (forecast (m3, h = 6), window(Xt, start = c(2018, 1),
end = c(2018, 6)))
m4 <- Arima(Yt, order = c(2, 0, 2), seasonal = c(0, 1, 1),
include.drift = TRUE)
accuracy (forecast (m4, h = 6), window (Xt, start = c(2018, 1),
end = c(2018, 6)))
m5 <- Arima(Yt, order = c(1, 0, 1), seasonal = c(2, 1, 1),
include.drift = TRUE)
accuracy (forecast (m5, h = 6), window(Xt, start = c(2018, 1),
end = c(2018, 6)))
m6 <- Arima(Yt, order = c(1, 1, 1), seasonal = c(0, 1, 1))
accuracy (forecast (m6, h = 6), window(Xt, start = c(2018, 1),
end = c(2018, 6)))

```

```

m7 <- Arima(Yt, order = c(1, 0, 1), seasonal = c(2, 1, 1),
            include.drift = TRUE, lambda = lambda)
accuracy (forecast (m7, h = 6), window(Xt, start = c(2018, 1),
            end = c(2018, 6)))
m8 <- Arima(Yt, order = c(1, 1, 1), seasonal = c(0, 1, 1),
            lambda = lambda)
accuracy (forecast (m8, h = 6), window(Xt, start = c(2018, 1),
            end = c(2018, 6)))

```

3. EVALUACIÓN DEL MODELO

```

tsdisplay(residuals(m1), main = "")
Box.test(residuals(m1), lag = 12, type = "Ljung", fitdf = 4)
Box.test(residuals(m1), lag = 24, type = "Ljung", fitdf = 4)
Box.test(residuals(m1), lag = 36, type = "Ljung", fitdf = 4)
Box.test(residuals(m1), lag = 48, type = "Ljung", fitdf = 4)
tsdiag(m1, 36)

```

4. REALIZACIÓN DE PRONOSTICOS CON EL MODELO

```

plot(forecast(m1, h = 6),
     ylim = c(min(Xt), max(Xt)),
     main = "",
     xlab = "Año",
     ylab = "Turistas Extranjeros",
     xaxt = "n")
axis(1, seq(2002, 2018, by = 1))
lines(Xt)
lines(fitted(m1), col = "tomato")
legend("topleft", c("Real", "Ajuste", "Pronóstico"),
     col = c("black", "tomato", "blue"), lty = 1)

```

- ANEXO 2 -

MODELOS DE REDES NEURONALES

```
rn1 <- nnetar(Yt, p = 13, P = 1, repeats = 1000, maxit = 300)
accuracy(forecast(rn1, h = 6), window(Xt, start = c(2018, 1),
  end = c(2018, 6)))
rn2 <- nnetar(Yt, p = 13, P = 2, repeats = 1000, maxit = 300)
accuracy(forecast(rn2, h = 6), window(Xt, start = c(2018, 1),
  end = c(2018, 6)))
rn3 <- nnetar(Yt, p = 2, P = 1, repeats = 1000, maxit = 300)
accuracy(forecast(rn3, h = 6), window(Xt, start = c(2018, 1),
  end = c(2018, 6)))
rn4 <- nnetar(Yt, p = 2, P = 2, repeats = 1000, maxit = 300)
accuracy(forecast(rn4, h = 6), window(Xt, start = c(2018, 1),
  end = c(2018, 6)))
rn5 <- nnetar(Yt, p = 2, P = 2, size = 4, repeats = 1000, maxit = 300)
accuracy(forecast(rn5, h = 6), window(Xt, start = c(2018, 1),
  end = c(2018, 6)))
rn6 <- nnetar(Yt, p = 4, P = 1, repeats = 1000, maxit = 300)
accuracy(forecast(rn6, h = 6), window(Xt, start = c(2018, 1),
  end = c(2018, 6)))
rn7 <- nnetar(Yt, p = 4, P = 2, repeats = 1000, maxit = 300)
accuracy(forecast(rn7, h = 6), window(Xt, start = c(2018, 1),
  end = c(2018, 6)))
rn8 <- nnetar(Yt, p = 4, P = 2, size = 8, repeats = 1000, maxit = 300)
accuracy(forecast(rn8, h = 6), window(Xt, start = c(2018, 1),
  end = c(2018, 6)))
plot(forecast(rn7, h = 6),
  ylim = c(min(Xt), max(Xt)),
  main = "",
```

```
xlab = "Año",  
ylab = "Turistas Extranjeros",  
xaxt = "n")  
axis(1, seq(2002, 2018, by = 1))  
lines(Xt)  
lines(fitted(rn7), col = "tomato")  
legend("topleft", c("Real", "Ajuste", "Pronóstico"),  
      col = c("black", "tomato", "blue"), lty = 1)
```

- ANEXO 3 -

Serie mensual de los visitantes extranjeros al Santuario Histórico de Machupicchu.

Año	Meses											
	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Sep	Oct	Nov	Dic
2002	23693	18299	23670	21761	25478	24215	32263	38591	26335	29146	25345	18193
2003	21401	17171	22129	25637	25259	26686	37100	42518	31302	32598	29884	22120
2004	25464	21046	26703	31868	33057	33222	44509	49160	36343	41265	35325	26537
2005	33642	25952	38939	33133	40477	39966	51515	57522	42633	44401	39907	27228
2006	35820	28113	38142	40187	41220	36940	50288	54706	42784	45532	39645	27364
2007	40860	31593	42319	44401	44840	43595	59528	63256	46337	51576	46255	33608
2008	48061	36025	50798	45454	52340	50041	67896	68085	55529	57904	45550	38432
2009	48937	36903	45105	52429	48300	47373	62408	59941	51132	52036	45606	31710
2010	44674	37458	48446	38545	45459	44890	62083	59333	47598	51373	44039	34748
2011	49223	36181	46374	52608	54594	55279	76546	73009	62357	65026	56909	42853
2012	60342	42000	54003	67211	69378	62275	80931	72930	68670	73025	62392	49312
2013	61600	41556	62578	64929	77167	67473	84452	79980	73927	76345	63875	50466
2014	61878	38485	62231	74005	83047	68815	86246	84648	78906	80554	66647	56729
2015	68057	48533	70048	78489	87432	76040	99848	90450	83619	79216	70183	59138
2016	69652	56236	81928	72931	96398	87529	107363	104298	94052	91218	71026	64133
2017	78711	59561	73100	90879	98759	94914	113817	111229	98484	98022	82918	70290
2018	84898	63051	93276	102112	119363	108128						

Fuente: Ministerio de Cultura – Dirección Desconcentrada de Cultura