



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



**MODELO UNIVARIANTE PARA DESCRIBIR Y PRONOSTICAR**  
**EL COMPORTAMIENTO DE LA SERIE HISTÓRICA DE**  
**FEMINICIDIOS DEL PERÚ, PERIODO 2010 - 2021**

**TESIS**

**PRESENTADA POR:**

**Bach. GABRIELA YAMILET BALCONA CCAMA**

**Bach. LEIDY MADU LOPEZ MAMANI**

**PARA OPTAR EL TÍTULO PROFESIONAL DE:**  
**INGENIERO ESTADÍSTICO E INFORMÁTICO**

**PUNO – PERÚ**

**2024**



## Reporte de similitud

NOMBRE DEL TRABAJO

**MODELO UNIVARIANTE PARA DESCRIBIR Y PRONOSTICAR EL COMPORTAMIENTO DE LA SERIE HISTÓRICA DE FEMINICIDIOS DEL PERÚ, PERIODO 2010 - 2021**

AUTOR

**GABRIELA YAMILET BALCONA CCAMA  
Y LEIDY MADU LOPEZ MAMANI**

RECuento DE PALABRAS

**24947 Words**

RECuento DE CARACTERES

**137938 Characters**

RECuento DE PÁGINAS

**132 Pages**

TAMAÑO DEL ARCHIVO

**2.4MB**

FECHA DE ENTREGA

**Jul 16, 2024 8:07 PM GMT-5**

FECHA DEL INFORME

**Jul 16, 2024 8:09 PM GMT-5**

### ● 12% de similitud general

El total combinado de todas las coincidencias, incluidas las fuentes superpuestas, para cada base de datos.

- 10% Base de datos de Internet
- Base de datos de Crossref
- 8% Base de datos de trabajos entregados
- 2% Base de datos de publicaciones
- Base de datos de contenido publicado de Crossref

### ● Excluir del Reporte de Similitud

- Material bibliográfico
- Material citado
- Material citado
- Coincidencia baja (menos de 12 palabras)



Firmado digitalmente por JUAREZ  
VARDAS Juan Carlos FAU  
20145496170.asf  
Módulo: Soy el autor del documento  
Fecha: 16.07.2024 20:54:16 -05:00



Firmado digitalmente por MJATA  
PANCA Petroy FAU 20145496170.asf  
Módulo: Soy el autor del documento  
Fecha: 16.07.2024 20:28:09 -05:00

Resumen



## DEDICATORIA

A Dios, fuente de toda sabiduría y fortaleza agradezco profundamente su guía y protección a lo largo de mi vida.

A mis queridos padres, Julio Cesar Balcona Ramos y Marleny Magdalena Ccama Ccama, quienes han sido mi mayor fuente de inspiración, les agradezco profundamente todo lo que soy como persona. Gracias papitos por todo el apoyo y amor incondicional que me brindaron durante el transcurso de mi carrera profesional y mi vida. Sin ustedes este logro no hubiera sido posible, por lo que dedico esta tesis a ustedes con todo mi corazón, los amo hasta el infinito y más allá.

A mi hermana y hermanos, Valeria, André y Brayan quienes con su alegría y compañía hicieron de este camino una experiencia más llevadera, divertida y gratificante. Gracias por ser mis cómplices sus risas me han dado la fuerza para superar cada obstáculo, ustedes me han demostrado el verdadero significado de lo que es familia.

A David Oscar Cueva Condori, por estar a mi lado en cada paso, brindándome su cariño, comprensión y ánimo en los momentos más difíciles. Eres mi roca y mi refugio, la persona en la que siempre puedo confiar. Gracias por haber llenado mis días de motivación, dulzura y alegría.

**Gabriela Yamilet. Balcona Ccama.**



## DEDICATORIA

A Dios, por ser mi guía y fortaleza en cada paso de este camino y por brindarme la sabiduría para superar cada desafío.

A mi amada madre, Carmen Candelaria Mamani Ventura, tus sacrificios, tu amor sin límites y tu incansable esfuerzo han sido la fuerza motriz que me impulso a lo largo de este viaje. Desde que tengo memoria, has sido mi ejemplo de fortaleza y perseverancia. Cada logro que alcanzo es una muestra de tus enseñanzas y de tu infinito amor. Gracias mamita por ser mi pilar, mi guía y mi mayor inspiración.

A mi querida hermana Yenelyn Yessenya, tu tenacidad y lucha han hecho que seas mi modelo a seguir. Gracias por ser mi confidente, mi mejor amiga, mi fuente constante de fortaleza en los momentos más difíciles y por brindarme tu amor incondicional que siempre me ha sostenido.

A mi sobrinito Adrián, eres la luz más brillante en nuestras vidas, desde tu llegada has llenado de alegría y felicidad cada día de mi vida.

A Erick Yair, gracias por ser mi faro en los momentos de incertidumbre, por tu apoyo incondicional y cariño que han sido el pilar de mi fortaleza. Tus palabras de aliento y tu fe inquebrantable en mí han sido la chispa que ha encendido mi determinación. Gracias por ser mi inspiración y motivación constante.

**Leidy Madu Lopez Mamani.**



## AGRADECIMIENTOS

A cada una de nuestras familias por su amor, apoyo incondicional y comprensión. Su confianza y apoyo en nosotras ha sido una fuente inagotable de motivación, energía y fortaleza.

A nuestros docentes de Estadística e Informática, quienes nos guiaron con su vasto conocimiento y experiencia. Su dedicación y compromiso fueron fundamentales para la culminación de este trabajo. En especial, queremos agradecer al Dr. Percy Huata Panca, quién nos ha orientado y motivado en cada etapa de nuestra investigación, así mismo a nuestros jurados, a la Dra. Teresa Pola Álvarez Rozas, Dr. Samuel Donato Pérez Quispe y M. Sc. Edgardo Quispe Yapó por su orientación, paciencia y valiosos consejos a lo largo de todo el proceso de investigación los cuales han sido fundamentales para la culminación de esta tesis.

Finalmente, a la Universidad Nacional del Altiplano y a todos los profesores y personal administrativo, por proporcionar un ambiente académico de excelencia y por su dedicación al desarrollo de sus estudiantes.

¡A todos que fueron parte de esta investigación, mil gracias!

**Gabriela Yamilet Balcona Ccama.**

**Leidy Madu Lopez Mamani.**



# ÍNDICE GENERAL

	Pág.
<b>DEDICATORIA</b>	
<b>AGRADECIMIENTOS</b>	
<b>ÍNDICE GENERAL</b>	
<b>ÍNDICE DE TABLAS</b>	
<b>ÍNDICE DE FIGURAS</b>	
<b>ACRÓNIMOS</b>	
<b>RESUMEN .....</b>	<b>14</b>
<b>ABSTRACT.....</b>	<b>15</b>
<b>CAPÍTULO I</b>	
<b>INTRODUCCIÓN</b>	
<b>1.1. PLANTEAMIENTO DEL PROBLEMA.....</b>	<b>18</b>
<b>1.2. FORMULACIÓN DEL PROBLEMA .....</b>	<b>19</b>
1.2.1. Problema general .....	19
1.2.2. Problemas específicos.....	19
<b>1.3. OBJETIVOS DE LA INVESTIGACIÓN.....</b>	<b>20</b>
1.3.1. Objetivo general.....	20
1.3.2. Objetivos específicos .....	20
<b>1.4. HIPÓTESIS DE LA INVESTIGACIÓN .....</b>	<b>20</b>
1.4.1. Hipótesis general.....	20
1.4.2. Hipótesis específicas.....	20



**1.5. JUSTIFICACIÓN DEL ESTUDIO..... 21**

**CAPÍTULO II**

**REVISIÓN DE LITERATURA**

**2.1. ANTECEDENTES DE LA INVESTIGACIÓN..... 23**

2.1.1. Antecedentes Internacionales ..... 23

2.1.2. Antecedentes Nacionales ..... 24

2.1.3. Antecedentes Locales ..... 27

**2.2. MARCO TEÓRICO ..... 29**

2.2.1. Pronóstico ..... 29

2.2.2. Series de tiempo..... 30

2.2.3. Componentes en las Series Temporales..... 30

2.2.4. Clasificación Descriptiva de las Series Temporales ..... 33

2.2.5. Modelo de composición de las componentes de una serie Temporal..... 34

2.2.6. Modelo ..... 36

2.2.7. Modelo Series de Tiempo ..... 37

2.2.8. Modelo Univariante ..... 37

2.2.9. Procesos Estocásticos ..... 38

2.2.10. Ruido Blanco ..... 42

2.2.11. Coeficiente de Correlación ..... 44

2.2.12. Autocorrelación ..... 45

2.2.13. Evaluación de Modelos AR, MA, ARMA y ARIMA ..... 49

2.2.14. Modelo Lineal Estacionario ..... 50



2.2.15.	Metodología Box – Jenkins .....	66
2.2.16.	Test de Shapiro – Wilk .....	70
2.2.17.	Test de Dickey – Fuller Aumentado (ADF) .....	71
2.2.18.	Estadístico de Box – Pierce .....	73
2.2.19.	Criterio de Información de Akaike (AIC): .....	74
2.2.20.	Criterio de información Bayesiano (BIC) .....	75
<b>2.3.</b>	<b>MARCO CONCEPTUAL .....</b>	<b>76</b>
2.3.1.	Mortalidad.....	76
2.3.2.	Mortalidad por sexo .....	76
2.3.3.	Violencia de género y violencia contra la mujer .....	77
2.3.4.	Política penal del delito de feminicidio .....	79
2.3.5.	Feminicidio .....	80
2.3.6.	Características del Feminicidio.....	81
2.3.7.	Causas del Feminicidio .....	81
2.3.8.	Consecuencias del Feminicidio .....	85
2.3.9.	Factores determinantes del feminicidio .....	86

### **CAPÍTULO III**

#### **MATERIALES Y MÉTODOS**

<b>3.1.</b>	<b>UBICACIÓN GEOGRÁFICA DEL ESTUDIO.....</b>	<b>89</b>
<b>3.2.</b>	<b>PERIODO DE DURACIÓN DE ESTUDIO.....</b>	<b>89</b>
<b>3.3.</b>	<b>PROCEDENCIA DEL MATERIAL UTILIZADO.....</b>	<b>89</b>
<b>3.4.</b>	<b>POBLACIÓN Y MUESTRA DE ESTUDIO .....</b>	<b>89</b>





3.4.1. Población .....	89
3.4.2. Muestra .....	90
<b>3.5. DISEÑO METODOLÓGICO.....</b>	<b>90</b>
3.5.1. Diseño de la Investigación .....	90
3.5.2. Enfoque de Investigación.....	90
3.5.3. Tipo de Investigación.....	91
3.5.4. Método de Investigación.....	91
<b>3.6. PROCEDIMIENTO DE RECOLECCIÓN DE DATOS .....</b>	<b>93</b>
3.6.1. Técnicas de recolección de datos.....	93
<b>3.7. VARIABLES .....</b>	<b>94</b>
<b>CAPÍTULO IV</b>	
<b>RESULTADOS Y DISCUSIÓN</b>	
<b>4.1. IDENTIFICACIÓN DEL MODELO ARIMA.....</b>	<b>96</b>
<b>4.2. ESTIMACIÓN DEL MODELO ARIMA .....</b>	<b>104</b>
<b>4.3. VALIDACIÓN DEL MODELO ARIMA .....</b>	<b>106</b>
<b>4.4. PRONÓSTICO.....</b>	<b>110</b>
<b>4.5. DISCUSIÓN .....</b>	<b>112</b>
<b>V. CONCLUSIONES.....</b>	<b>115</b>
<b>VI. RECOMENDACIONES .....</b>	<b>116</b>
<b>VII. REFERENCIAS BIBLIOGRÁFICAS.....</b>	<b>117</b>
<b>ANEXOS.....</b>	<b>125</b>



**ÁREA:** Análisis Univariado

**TEMA:** Modelo univariante para describir y pronosticar el comportamiento de la serie  
Histórica de Femicidios del Perú, periodo 2010 - 2021.

**FECHA DE SUSTENTACIÓN:** 17 de julio de 2024



## ÍNDICE DE TABLAS

	<b>Pág.</b>
<b>Tabla 1</b> Correlación de la Población y de la Muestra .....	44
<b>Tabla 2</b> Operacionalización de variables .....	94
<b>Tabla 3</b> Datos De Información .....	96
<b>Tabla 4</b> Estadísticas descriptivas de la serie casos de feminicidio en el Perú 2010-2021.....	98
<b>Tabla 5</b> Test de Shapiro-Wilk .....	100
<b>Tabla 6</b> Test de Dickey Fuller .....	101
<b>Tabla 7</b> Posibles modelos ARIMA .....	104
<b>Tabla 8</b> Estimación de parámetros del modelo $ARIMA(0,0,1) \times 1,0,112$ .....	105
<b>Tabla 9</b> Test de Box-Ljung .....	108
<b>Tabla 10</b> Métricas de precisión de los modelos ARIMA .....	109
<b>Tabla 11</b> Valores pronosticados casos de feminicidio en el Perú 2022 .....	111



## ÍNDICE DE FIGURAS

	<b>Pág.</b>
<b>Figura 1</b> Componentes en Series Temporales.....	31
<b>Figura 2</b> Funciones de Autocorrelación .....	46
<b>Figura 3</b> AR(1) con $\alpha > 1$ .....	52
<b>Figura 4</b> AR(1) con $\alpha < -1$ .....	53
<b>Figura 5</b> Procesos autorregresivos de orden 1 .....	54
<b>Figura 6</b> Procesos autorregresivos de orden 2 .....	55
<b>Figura 7</b> Procesos Medias Móviles de orden1:MA(1).....	60
<b>Figura 8</b> Series temporales económicas.....	64
<b>Figura 9</b> Metodología Box – Jenkins .....	67
<b>Figura 10</b> Diagrama de flujo de la metodología Box – Jenkins.....	92
<b>Figura 11</b> Gráfico de la serie casos de feminicidio 2010-2021 .....	97
<b>Figura 12</b> Descomposición multiplicativa de la serie casos de feminicidios 2010+- 2021 .....	98
<b>Figura 13</b> Histograma de la serie cosos de feminicidios 2010-2021 .....	99
<b>Figura 14</b> Función De Autocorrelación Simple y Parcial de la serie original. ....	102
<b>Figura 15</b> Distribución normal de los residuos del modelo seleccionado .....	107
<b>Figura 16</b> Función De Autocorrelación Simple Y Parcial de residuos del modelo seleccionado .....	108
<b>Figura 17</b> Serie original con respecto a la serie ajustada del modelo $ARIMA(0,0,1) \times$ $1,0,112$ .....	110
<b>Figura 18</b> Pronostico de casos de feminicidios en el Perú 2022.....	112



## ACRÓNIMOS

AURORA	: Programa Nacional para la Prevención y Erradicación de la Violencia Contra las Mujeres e Integrantes del Grupo Familiar.
CEM	: Centro de Emergencia Mujer.
CEIC	: Comité Estadístico Interinstitucional de la Criminalidad.
INEI	: Instituto Nacional de Estadística e Informática.
MININTER	: Ministerio del Interior.
OMS	: Organización Mundial de la Salud.
ONU	: Organización de las Naciones Unidas.
RAE	: Real Academia Española.
PNP	: Policía Nacional del Perú.
MINJUSDH	: Ministerio de Justicia y Derechos Humanos



## RESUMEN

La siguiente investigación es analítico- aplicada se realizó en la Región de Puno - Perú, con el propósito de evaluar las futuras proyecciones del Femicidio en el Perú. El femicidio es una preocupación global, y Perú, dentro de Latinoamérica, se sitúa en el tercio inferior en términos de tasa de femicidio. Por tanto, es crucial estudiar las proyecciones futuras de este fenómeno para implementar medidas preventivas y reducir el aumento de muertes. En este contexto, el objetivo principal es identificar un modelo de predicción mensual efectivo que se ajuste adecuadamente a la serie histórica. Esto permitirá optimizar la toma de decisiones a corto plazo en términos de asignación de recursos humanos, equipamiento y tecnología. La población del estudio abarcó el total de femicidios registrados desde el año 2001 hasta la actualidad según el Programa Nacional AURORA. La muestra se compuso de los datos recopilados por el programa desde el año 2010 hasta el año 2021. La metodología se fundamentó en el análisis de las propiedades probabilísticas de las series temporales. El resultado que se llegó es la estimación del mejor modelo univariante para predecir la serie original de femicidios en el Perú es un modelo  $Y_t = 10.16817 + 0.15179\varepsilon_{(t-1)} + 0.54948\varepsilon_{(t-12)} + 0.77771\varepsilon_{(t-24)} + \varepsilon_t$ . Se ha determinado que el modelo univariante más adecuado, el cual proporciona el mejor ajuste a la serie histórica mensual de femicidios en el Perú durante 2010-2021 cumple con todos los requisitos estipulados por la metodología Box-Jenkins.

**Palabras clave:** Femicidio, Estimación, Modelo, Parámetros, Predicción Mensual.



## ABSTRACT

This study is of analytical-applied nature and was carried out in the Puno Region of Peru in 2022, with the purpose of evaluating future projections of Femicide in Peru. Femicide is a global concern, and Peru, within Latin America, ranks in the bottom third in terms of femicide rate. Therefore, it is crucial to study future projections of this phenomenon to implement preventive measures and reduce the increase in deaths. In this context, the main objective is to identify an effective monthly prediction model that fits the historical series, thus facilitating short-term decision making regarding human resources, equipment and technology. The study population included all femicides recorded from 2001 to the present according to the National Program AURORA. The sample consisted of data collected by the program from 2010 to 2021. The methodology relied on analyzing the probabilistic characteristics of the time series. The result that was reached is the estimation of the best univariate model to predict the original series of femicides in Peru, a model  $Y_t = 10.16817 + 0.15179\varepsilon_{(t-1)} + 0.54948\varepsilon_{(t-12)} + 0.77771\varepsilon_{(t-12)} + \varepsilon_t$ . It has been concluded that the best univariate model that offers an optimal fit to the monthly historical series of femicides in Peru during 2010-2021 meets all the requirements stipulated by the Box-Jenkins methodology.

**Keywords:** Femicide, Estimation, Model, Parameters, Monthly Prediction.



# CAPÍTULO I

## INTRODUCCIÓN

La violencia que se ha dado contra las mujeres actualmente es reconocida como una seria infracción a los derechos humanos y un gran impedimento en el camino hacia la igualdad de género, que afecta a todas las mujeres (OMS, 2005). Una forma común de violencia contra las mujeres es la violencia de pareja, perpetrada por el marido o alguien cercano a la víctima (Heise y García, s. f.). Es indiscutible que a nivel mundial una parte considerable de mujeres están pasando por la violencia generada por sus parejas, una de cada tres mujeres ha sido objeto de violencia física, psicológica o sexual por parte de su pareja en algún momento de su vida (World Health Organization, 2013)

Las tasas de maltrato doméstico y feminicidio son especialmente elevadas en América Latina, una región con la segunda tasa de homicidio femenino más alta del mundo, solo superada por África World Health Organization (2013). Entre los 25 países que presentan tasas de feminicidio altas y muy altas, 14 se encuentran en América Latina y el Caribe Alvazzi del Frate (2011), esto puede llegar a explicar la razón por la que varios países han promulgado leyes que tipifican el feminicidio como un delito y se definen diversas formas de sanciones. Uno de los desafíos fundamentales al tratar el feminicidio es la falta de datos oficiales, actuales y fiables. Y el gran problema de comparar países, ya que los métodos de recopilación de datos difieren entre ellos.(Carrigan, s. f.)

El feminicidio, un fenómeno común en Perú, afecta a numerosas mujeres, sin importar su posición social. Estos crímenes ocurren en todos los niveles de la sociedad y traspasan las barreras políticas y económicas. Desafortunadamente, estos incidentes se presentan en todas las sociedades del mundo, sin distinción de situación económica, raza o cultura. No obstante, la diferencia radica en que ciertos estratos sociales denuncian más





que otros. La violencia contra las mujeres en la sociedad actual es una pandemia que afecta gravemente a la sociedad, ya que vulnera los derechos esenciales de las personas, pone en peligro sus vidas y obstaculiza el progreso hacia una sociedad inclusiva y democrática.

Las estadísticas son esenciales para comprender la realidad práctica, ya que influyen directamente en aspectos sociales. Su utilidad práctica suele ser más evidente que la de otras ciencias exactas, como las matemáticas (Batanero, 2001).

La mayoría de los métodos univariados de pronóstico se basan en datos estables y predecibles. Sin embargo, el procedimiento de Box-Jenkins introduce un componente subjetivo que permite seleccionar entre varios modelos y evaluar la calidad de las predicciones. Esto simplifica el proceso de tomar decisiones y desarrollar políticas. Por tales motivos es relevante, identificar el patrón de datos a lo largo del tiempo de feminicidios en el Perú, ya que esto nos asistirá en la toma de decisiones para prevenir y encontrar soluciones para reducir estos datos que son preocupantes por eso, el objetivo de este estudio es establecer los pronósticos utilizando un modelo univariado de la serie de feminicidios en el Perú. Para ello utilizaremos a la metodología Box-Jenkins, con información de los feminicidios ocurridos en el país por meses, durante el periodo 2010 – 2021, buscando luego un modelo adecuado que permitirá ajustar el comportamiento de la serie de Feminicidios

El presente estudio se estructura en cinco capítulos. En el Capítulo I, se presenta el planteamiento del problema, los objetivos, las hipótesis y la justificación que sustenta la realización de esta investigación. Seguidamente en el capítulo II, se detallan los antecedentes, así como el marco teórico y conceptual. Asimismo, en el Capítulo III, se describen los materiales y métodos empleados en el estudio, incluyendo el lugar donde



se llevó a cabo la investigación, la población objeto de estudio, entre otros aspectos relevantes. Finalmente, en el Capítulo IV, se muestran los resultados hallados y sus respectivas conclusiones y recomendaciones.

### **1.1. PLANTEAMIENTO DEL PROBLEMA**

La violencia familiar y sexual constituye un grave problema en el Perú, el cual es frecuente en el que se refiere a las tasas de enfermedades que impactan sistemáticamente a la población, en particular a mujeres, niñas y niños. Esto representa una transgresión a los derechos fundamentales de los peruanos y un obstáculo importante para el avance humano y social (WORLD BANK GROUP, 2021).

El feminicidio ya no es un tema ignorado y ahora se discute en todos los estratos sociales. Tanto las autoridades públicas como la sociedad en general están reconociendo el feminicidio como un grave problema social que necesita ser abordado. (Soto & Cenitagoya Garín, 2017)

El abuso psicológico y físico contra las mujeres causa un impacto directo en la disminución de su autoestima, así como en su salud mental y física, manifestándose en depresión, ansiedad, sentimientos de culpa y rechazo, entre otros. Además, destruye el concepto de familia como un lugar seguro y protector, limita el desarrollo de habilidades humanas y conduce a una mayor pobreza debido al deterioro del capital humano. Esta violencia social desenfrenada, si no se controla, puede culminar en feminicidio (Viviano Llave, 2007)

La función principal de la estadística radica en la recopilación y organización de datos de diversas índoles, con el fin de generar informes estadísticos que proporcionen comprensión sobre una amplia gama de temas, como el feminicidio en el Perú. Es crucial destacar esto, ya que la estadística se transforma en una disciplina que no solo trata de



números, sino que nos ofrece una visión directa de las realidades sociales.(Villegas Zamora, 2019)

Finalmente, es esencial emplear la metodología de Box-Jenkins para analizar el patrón de la serie temporal de feminicidios en el Perú, se pretende identificar un modelo adecuado que permita ajustar el comportamiento de esta serie, lo cual es crucial para tomar decisiones informadas.

## **1.2. FORMULACIÓN DEL PROBLEMA**

A partir de los argumentos mostrados en el planteamiento del problema, se han formulado las siguientes interrogantes de investigación:

### **1.2.1. Problema general**

¿Cuál es el tipo de modelo univariante que mejor se ajusta a la serie histórica mensual de Feminicidios del Perú, durante el periodo 2010 hasta el 2021 para describir y pronosticar el comportamiento de la serie histórica?

### **1.2.2. Problemas específicos**

- ¿Cuál es el mejor modelo identificado que se ajusta a la serie histórica para pronosticar y describir el comportamiento del Feminicidio en el Perú, durante el periodo 2010 – 2021?
- ¿Cuáles son los pronósticos alcanzados para validar el número de Feminicidios en el Perú usando el modelo adecuado?



### **1.3. OBJETIVOS DE LA INVESTIGACIÓN**

#### **1.3.1. Objetivo general**

Determinar el tipo de modelo univariante que mejor se ajusta a la serie histórica mensual de Femicidios del Perú, durante el periodo 2010 hasta el 2021 para describir y pronosticar el comportamiento de la serie histórica.

#### **1.3.2. Objetivos específicos**

- Identificar, estimar y validar el mejor modelo que se ajusta a la serie histórica para pronosticar y describir el comportamiento del Femicidio en el Perú, durante el periodo 2010 – 2021.
- Obtener Pronósticos validados del número de Femicidios en el Perú usando el modelo adecuado.

### **1.4. HIPÓTESIS DE LA INVESTIGACIÓN**

#### **1.4.1. Hipótesis general**

El modelo univariante que se ajusta mejor a la serie histórica mensual Femicidios en el Perú para pronosticar y describir su comportamiento, es un modelo multiplicativo ARIMA  $(p*d*q)$   $(P*D*Q)$ , constituido por una parte regular y otra estacional.

#### **1.4.2. Hipótesis específicas**

- El modelo determinado mediante la metodología de Box-Jenkins, se ajusta a la serie histórica para pronosticar y describir el comportamiento de la serie de Femicidios del Perú, periodo 2010 -2021.



- Los pronósticos alcanzados para validar el número de Femicidios en el Perú usando el modelo adecuado será constante a lo largo del año 2022.

## 1.5. JUSTIFICACIÓN DEL ESTUDIO

Durante el año 2021, se registraron 163,797 casos de violencia contra mujeres y miembros de sus familias, de los cuales 140,833 (86%) afectaron a mujeres (AURORA, 2021)

Determinado por Diana Russell como “El asesinato de una mujer específicamente por el hecho de ser mujer”, el feminicidio no solo se limita a los asesinatos la violencia contra las mujeres perpetrada por hombres debido a su género, existe también una minoría significativa de asesinatos la violencia contra mujeres perpetrada por otras mujeres debido a su género (PATH et al., 2008).

El incremento de los homicidios de mujeres a escala global evidencia la prevalencia de la violencia contra las mujeres como un problema social generalizado. Este fenómeno se origina en las disparidades de género en las relaciones y en la aceptación social de este tipo de violencia.

En Perú, se han conseguido progresos destacables en esta área legal y político para prevenir, condenar y erradicar el feminicidio. Sin embargo, en muchas ocasiones, estas medidas no se aplican en la práctica. Cuando las víctimas denuncian maltratos y agresiones a su integridad, las autoridades a menudo toman medidas insuficientes, lo que lamentablemente resulta en más mujeres que son asesinadas.

El propósito de esta investigación es aportar al apoyo de las autoridades competentes puedan disminuir el número de muertes a causa del Femicidio. Para ello,



se empleará el método de series temporales con el enfoque de Box-Jenkins, con el fin de comprender y analizar el patrón de ocurrencia de Femicidios en el Perú.



## CAPÍTULO II

### REVISIÓN DE LITERATURA

En esta etapa se presenta el marco teórico relacionado con el tema de investigación, seguido de una sección adicional que revisa los antecedentes de estudios previos.

#### 2.1. ANTECEDENTES DE LA INVESTIGACIÓN

##### 2.1.1. Antecedentes Internacionales

Coutin (2007) Menciona que la capacidad de tomar medidas oportunas permite a las autoridades mitigar las consecuencias, algunas veces letales, de enfermedades y otros daños. El ARIMA, desarrollado por Box y Jenkins en 1976, ha ganado popularidad en la salud pública a nivel universal gracias a su capacidad para modelar con precisión la evolución de eventos de salud y anticipar el curso de enfermedades transmisibles bajo vigilancia. Este estudio ofrece una revisión metodológica y ejemplos prácticos que muestran cómo se emplea el ARIMA para realizar pronósticos. Finalmente, en sus resultados mostró la predicción de la atención médica semanal para enfermedades diarreicas agudas e infecciones respiratorias agudas en Cuba entre 1998 y 2004. Además, se comparan los valores pronosticados con los observados, con lo cual pudo concluir que estos modelos son estables y apropiados para prever estos eventos de salud.

Mondal et al., (2014) El autor realizó un estudio sobre la eficacia del modelo de media móvil autorregresiva integrada (ARIMA) en cincuenta y seis valores de diferentes sectores. Escogió el modelo ARIMA por su sencillez y su amplia aceptación. También el investigador estudió el efecto sobre la precisión de



las predicciones considerando múltiples conjuntos de datos de períodos anteriores. La comparación y ajuste de parámetros del modelo ARIMA llevó a cabo utilizando el criterio de información de Akaike (AIC). Este enfoque les permitió determinar la eficacia del modelo en predecir el comportamiento futuro de los eventos estudiados, proporcionando una medida objetiva de su desempeño predictivo.

Ariyo et al., (2014) Sostiene que la predicción del precio de las acciones es un tema importante en finanzas y economía que ha despertado el interés de los investigadores a lo largo de los años por desarrollar mejores modelos de predicción. Menciona que los modelos de media móvil autorregresiva integrada (ARIMA) se han explorado en la literatura para la predicción de series temporales. Por lo que presenta un amplio proceso de construcción de un modelo predictivo del precio de las acciones utilizando el modelo ARIMA. Empleando datos bursátiles de la Bolsa de Nueva York (NYSE) y de la Bolsa de Nigeria (NSE), por lo que el autor desarrolló un modelo de predicción del precio de las acciones. Los resultados obtenidos mostraron que el modelo ARIMA presenta un gran potencial para la predicción a corto plazo y puede disputar favorablemente con las técnicas existentes para la predicción del precio de las acciones.

### **2.1.2. Antecedentes Nacionales**

Carranza (2019) en su investigación denominada “Análisis estadístico de las exportaciones de bienes en la Comunidad Andina, Enero 2010 - Junio 2018”, empleó el análisis de series temporales utilizando la “Metodología de Box-Jenkins”, empleando datos obtenidos de la página del Sistema Integrado de Comercio Exterior (SICEXT) de la Comunidad Andina (CAN). Se determinó que





la serie identificada no era estacionaria, y se seleccionó un modelo de pronóstico SARIMA (1,1,0)(1,1,0)<sub>12</sub> sin intercepto. Se llevó a cabo una proyección para el período de enero a junio de 2018. La estimación de los pronósticos fue realizada utilizando el Porcentaje de Error Medio Absoluto (PEMA) se situó en un 4.27%, mientras que el Porcentaje Medio de Error (PME) fue de -0.04%. De esta manera, el autor determinó que el modelo no muestra sobreestimación, ni subestimación en las exportaciones mensuales en miles de dólares hacia el mundo dentro de la Comunidad Andina.

Alva (2017) el propósito de esta investigación fue “identificar el modelo más adecuado para comprender cómo varían un análisis de las defunciones mensuales en el Hospital Belén”, este estudio se extendió a lo largo del tiempo, ya que recopiló datos mensuales desde enero de 2011 hasta diciembre de 2015. La investigación empleó la metodología de Box Jenkins para identificar los siguientes modelos: un ARIMA (6, 0, 6) para el total de defunciones en los servicios de hospitalización, un ARIMA (13, 0, 13) para cardiología, un ARIMA (0, 0, 18) para cirugía general, un ARIMA (6, 0, 6) para medicina interna y un ARIMA (0, 0, 24) para pediatría. Llegó a concluir que los modelos ARIMA obtenidos en ese estudio muestran una similitud notable con los datos reales proporcionados por el Hospital Belén de Trujillo.

Acosta y Chapoñan (2022) buscaron “determinar el modelo de series temporales que mejor describa a la producción mensual de espárragos en el Perú durante el período de Enero de 2007 a Diciembre de 2018”. Utilizando una metodología observacional tanto retrospectiva como prospectiva, como parte de una investigación teórica, empírica y longitudinal, se analizó la obtención mensual de espárragos en miles de toneladas en el Perú entre los años 2007 y 2018,



considerando el periodo de enero a diciembre. Los resultados indicaron que el modelo SARIMA (1,1,1) (2,1,0) demostró una precisa capacidad de predicción en la producción de espárragos. Además, se encontró que los parámetros del modelo eran estadísticamente significativos y que los errores asociados se distribuían de la siguiente manera: un error cuadrático medio de 2.2035, un error absoluto medio de 1.5899, un error porcentual medio de -1.7234 y un error porcentual absoluto medio de 5.4501. En consecuencia, se concluyó que el método de Box Jenkins, aplicando el modelo SARIMA (1,1,1) (2,1,0), proporciona una predicción precisa de la producción de espárragos, ya que los residuos del modelo satisfacen con los supuestos de independencia y normalidad.

Ramos (2020) comparó la “capacidad predictiva entre el modelo SARIMA y el modelo de redes neuronales artificiales (RNA) basados en los ingresos tributarios mensuales del gobierno central del Perú”. Los datos se obtuvieron desde enero de 2003 hasta diciembre de 2018 a partir del sitio web del Banco Central de Reserva del Perú. La serie de ingresos tributarios presentó tanto estacionalidad como una tendencia ascendente lineal, las cuales se corrigieron al aplicar una diferencia simple y a su vez una diferencia estacional. Además, se identificó la existencia de una variación cíclica que abarcaba más de un año en la serie. Tras analizar las funciones de autocorrelación simple (ACF) y parcial (PACF) de la serie diferenciada, es por ello que se optó por los modelos SARIMA (2,1,0)(1,1,1) y NNAR(2,1,5).

Damián y Sandoval (2018) generó un modelo de pronóstico óptimo para el Índice Mensual de Producción de Electricidad en Perú durante el lapso de tiempo comprendido entre 2006 y 2015. Para ello el estudio clasificó como descriptivo predictivo, en el que utilizó un enfoque longitudinal retrospectivo. Los



datos utilizados fueron recabados del Instituto Nacional de Estadística e Informática (INEI). La metodología que usó fue Box-Jenkins, en el que analizó los datos, y llegó a concluir que el modelo SARIMA (1, 1, 0) (0, 1, 1)<sup>12</sup> proporciona la mejor explicación para el comportamiento del Índice Mensual de Producción de Electricidad durante el periodo mencionado.

### 2.1.3. Antecedentes Locales

Mamani (2020) En su investigación el propósito del estudio fue realizar una estimación de la demanda mensual de horas de maquinaria agrícola en la municipalidad de Caracoto para el año 2020, en el que usó el enfoque del modelo de series temporales SARIMA, en el que empleó la “metodología Box-Jenkins” y programó en el lenguaje R para llevar a cabo el análisis, llegó al resultado, identificó el modelo de predicción SARIMA (2,1,0) (1,1,1)<sup>12</sup>, el cual ofrece una mejor explicación de la serie de datos, finalmente llegó a concluir que este modelo puede realizar un pronóstico para el año 2020 con un nivel de confianza del 95%.

Juculaca (2019) tuvo como objetivo de poder “determinar el tipo de modelo univariante más adecuado para prever el número de casos de infecciones respiratorias agudas, neumonía y defunciones en niños menores de 5 años en la Dirección Regional de Salud Puno durante el año 2018”. El autor utilizó un diseño de investigación descriptivo y longitudinal para analizar las tendencias. empleó la “Metodología Box-Jenkins” para determinar el modelo más conveniente que se adaptara significativamente a los datos disponibles. Este proceso consistió en varias fases: análisis inicial de la serie, identificación del modelo más adecuado, estimación de los parámetros del modelo, comprobación y validación mediante el análisis de los residuos. Finalmente, la predicción de los modelos de pronóstico



más adecuados para las series correspondientes a los años 2018, 2019 y 2020, fueron los modelos SARIMA (2,0,0)(0,1,1) 12, SARIMA (1,0,1)(2,0,0)12 y SARIMA (2,0,1)(2,0,1) 12, los cuales describen las características de las series.

Melo (2016) tuvo como propósito anticipar las “proyecciones futuras del comportamiento de la mortalidad general intrahospitalaria en el Hospital Regional Manuel Núñez Butrón de Puno”, con el fin de planificar recursos humanos, infraestructura, equipamiento tecnológico y financiero. Entonces, el investigador se propuso hallar el modelo más preciso para predecir mensualmente a corto plazo utilizando la “Metodología Box-Jenkins” en la modelación de la serie temporal de muertes intrahospitalarias. Como resultado, identificó que el mejor modelo univariante para estas predicciones fue el SARIMA (2, 1, 0) (0, 1, 1) 12, el cual se ajustó de manera óptima a la serie inicial de mortalidad intrahospitalaria en el Hospital Regional Manuel Núñez Butrón - Puno.

López (2018) tenía la finalidad de “identificar los modelos univariantes más adecuados para pronosticar el número de casos de aborto y preeclampsia en el servicio de obstetricia hospitalaria del Hospital Regional Manuel Núñez Butrón de Puno”. El estudio se centró en mujeres atendidas en dicho servicio, cuyas edades oscilaban entre 15 y 47 años, y que habían sido diagnosticadas con aborto o preeclampsia. Para el análisis de los datos, aplicó la “metodología de Box-Jenkins”. Como resultado, el autor determinó que los modelos univariantes más adecuados para realizar predicciones fueron: ARIMA (1,0,1) (0,1,1) para el número de casos de aborto, y ARIMA (1,1,2) (0,1,1) para el número de casos de preeclampsia.



Machaca (2017) realizó “proyecciones sobre las variaciones de la desnutrición crónica en niños menores de 5 años en el Hospital Manuel Núñez Butrón de Puno”, en el que busca determinar el mejor modelo univariante utilizando la “metodología de Box-Jenkins”, llegó al resultado, que el modelo de predicción mensual ARIMA (0, 0, 1) (0, 0, 1)<sup>12</sup> el modelo mostró ser más adecuado para analizar y anticipar el patrón de desnutrición crónica en niños menores de 5 años. Se determinó que replicar la serie de manera precisa ofrecía una herramienta efectiva para describir y prever su comportamiento futuro.

## 2.2. MARCO TEÓRICO

### 2.2.1. Pronóstico

Un pronóstico se refiere a una estimación o predicción futura sobre el valor de una variable aleatoria o serie temporal, basada en datos históricos y técnicas de modelado estadístico. El objetivo principal de realizar pronósticos es anticipar el comportamiento futuro de un fenómeno o proceso, lo que puede contribuir en la toma de decisiones y la planificación estratégica.(Villareal, 2016)

Los pronósticos pueden aplicarse en una amplia gama de contextos, como en economía, finanzas, meteorología, ventas, inventarios, producción, entre otros.

Los métodos utilizados para realizar pronósticos pueden variar dependiendo del tipo de datos disponibles, la naturaleza del fenómeno que se está pronosticando y el horizonte temporal de interés. Algunas técnicas comunes incluyen modelos de series temporales (como ARIMA y modelos de suavizado exponencial), análisis de regresión, modelos de aprendizaje automático (como redes neuronales), entre otros.(Villareal, 2016)



### **2.2.2. Series de tiempo**

Según González (2009) una serie temporal se define como una secuencia organizada y cronológica de datos observados en un periodo específico, las cuales cada una está asociada a un momento en el tiempo. Las series de tiempo son usadas para estudiar el comportamiento de una o más variables, los cuales se encuentran en cualquier campo de la ciencia y de la Demografía, ya que se analizan y examinan conjuntos de datos como las series de Población Total y las Tasas de Natalidad para comprender mejor los patrones y tendencias relacionadas con la población a lo largo del tiempo, Femicidio, entre otras.

Dado que las series temporales presentan particularidades únicas, se crean modelos adaptados para capturar y aprovechar la relación entre las observaciones en secuencia, lo que facilita un análisis más detallado y una comprensión más clara de los datos en el transcurso del tiempo.

Por lo que los estudios de series temporales se refieren al conjunto de métodos utilizados para estudiar observaciones secuenciales que están interrelacionadas y ordenadas en el tiempo. Dentro de este análisis, se emplean herramientas como modelos que pueden simular y representar adecuadamente el comportamiento de la variable de interés a lo largo del tiempo. (Pérez López, 2008)

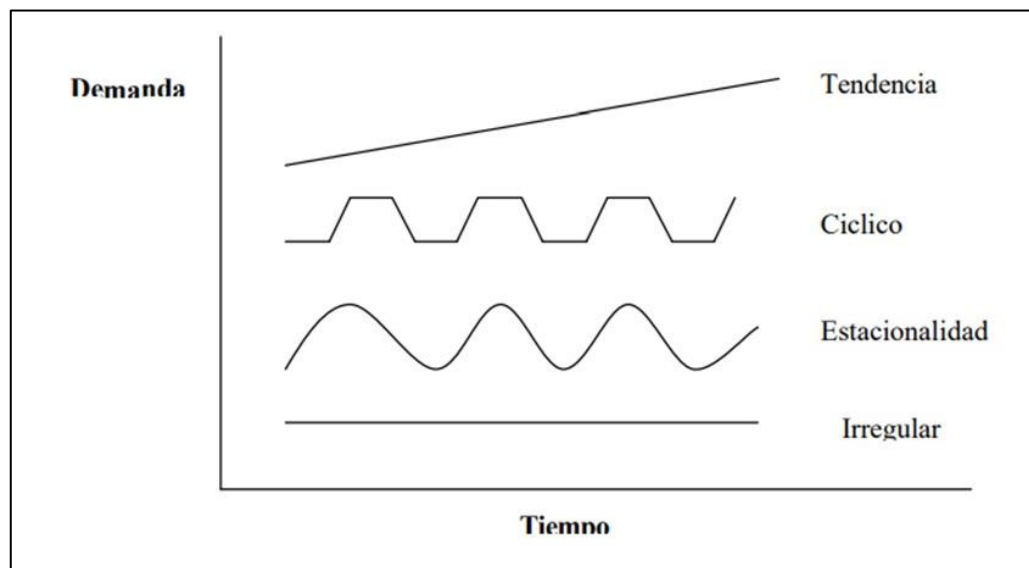
### **2.2.3. Componentes en las Series Temporales**

Según Chatfield (1987) los métodos habituales de análisis de series de tiempo, son los que se ocupan principalmente de descomponer la variación de una serie en: tendencia, estacionalidad, variación, otros cambios cíclicos y el resto de fluctuaciones irregulares. Este enfoque es especialmente valioso cuando la variación está dominada por la tendencia y/o la estacionalidad. Sin embargo, cabe

señalar que la descomposición no suele ser única a menos que se cumplan determinados supuestos. Estas técnicas descriptivas pueden implicar algún tipo de modelización, explícita o implícita, lo que demuestra la difusa frontera que separa las técnicas descriptivas de las inferenciales.

### Figura 1

#### *Componentes en Series Temporales*



Nota: (Hankhe & Reitsch, 1996).

#### 2.2.3.1. Tendencia

Una serie temporal con tendencia incluye un elemento de largo plazo que refleja el aumento o la disminución gradual de la serie a lo largo de un período específico. En otras palabras, una serie temporal exhibe una tendencia cuando su promedio cambia con el tiempo, lo que implica que se anticipa un aumento o una disminución a lo largo del período para el cual se está realizando el pronóstico (Hankhe y Reitsch, 1996)



### **2.2.3.2. Componente Cíclico o Variación Cíclica**

Según Caridad y Ocerin (2005) un componente cíclico son variaciones periódicas que se repiten varias veces, en general su amplitud es superior a un año, por lo que la estructura de estos ciclos son variables en función a la serie. Estas variaciones periódicas no suelen ser regulares, debido que presentan picos altos o bajos.

La variación cíclica consiste en un grupo de oscilaciones con estructura de ondas o ciclos que se extienden más allá de un año y son causadas por cambios. Esta variación representa la discrepancia entre los valores anticipados de una variable (la tendencia) y los valores observados, es decir, la fluctuación residual que rodea la tendencia (Hankhe & Reitsch, 1996).

### **2.2.3.3. Estacional**

De acuerdo con Abril (2011) numerosas series están afectadas por variables estacionales, como el clima, cuyos efectos pueden ser modelados mediante un componente periódico con un período fijo y conocido. La estacionalidad también puede ser considerada como una función determinística del tiempo, aunque, al igual que con la tendencia, podría ser vista como aleatoria. Por lo general, la parte determinística de la estacionalidad puede ser representada de dos maneras: mediante un conjunto de variables ficticias (llamadas "dummy") que suman cero durante el período estacional, o como una combinación de senos y cosenos con frecuencias que coinciden con el período estacional.





#### **2.2.3.4. Fluctuaciones Irregulares**

Una vez eliminadas la tendencia y las variaciones cíclicas de un conjunto de datos, queda un conjunto de residuos, la estacionalidad ha sido consistentemente considerada como una variable con ciertas propiedades aleatorias que se ajustan a las características específicas del fenómeno que genera la serie temporal en análisis. Por lo que se examina diversas técnicas de análisis de series de este tipo para ver si algunas de las variaciones aparentemente irregulares pueden explicarse en términos de modelos de probabilidad, como modelos de medias móviles o autorregresivos, por lo que alternativamente, se pueden ver si aún queda alguna variación cíclica en los residuos.(Chatfield, 1987)

Menciona Hankhe y Reitsch (1996) el comportamiento irregular abarca fluctuaciones derivadas de eventos impredecibles o no periódicos, como condiciones climáticas inusuales, conflictos laborales, situaciones de guerra, especulaciones sobre conflictos bélicos, procesos electorales y modificaciones en la legislación.

#### **2.2.4. Clasificación Descriptiva de las Series Temporales**

##### **2.2.4.1. Estacionarias**

Según Hankhe y Reitsch (1996) una serie estacionaria es aquella que se caracteriza por mantener sus propiedades estadísticas fundamentales, como la media y la varianza, las cuales son constantes a lo largo del tiempo. Esto implica que no experimenta un crecimiento o una disminución significativa en su tendencia central. Gráficamente, los valores en la serie tienden a mantenerse dentro de un rango definido por

una media y una varianza constantes. Por lo cual, se concluye que una serie estacionaria se caracteriza por mantener un valor promedio constante a lo largo del tiempo, sin experimentar cambios significativos en su tendencia central.

#### **2.2.4.2. No Estacionarias**

En este tipo de series cambian en el tiempo la tendencia o la variabilidad. Por lo que los cambios en la media llegan a determinar una tendencia positiva o negativa a largo plazo, es por ello, que la serie no oscila alrededor de un valor constante. Esto es a causa de que una observación en un extremo de la media general se orienta a ir seguida de un gran número de observaciones en el mismo lado de la media debido a la tendencia. Un correlograma de este tipo tiende a ser difícil de deducir, ya que la tendencia domina todas las demás características. De hecho, la función de autocorrelación de la muestra sólo tiene sentido para las series temporales estacionarias, por lo que debe eliminarse cualquier tendencia antes de calcularla. (Chatfield, 1987)

#### **2.2.5. Modelo de composición de las componentes de una serie Temporal**

Según Uriel (1985) los modelos de series temporales representan estructuras teóricas que pueden ser determinísticas, aleatorias o una combinación de ambas. Las cuales se emplean para examinar y comprender el comportamiento de una serie de tiempo en el transcurso del tiempo.

Variabes Temporales: Son aquellas variables que se observan durante un periodo en el tiempo, por lo que se simboliza como  $Y_t$



Y: es la variable

t: es el tiempo o momento

Serie Temporal: Son conjuntos de t observaciones. Por tanto, es una observación distinta en el tiempo en la misma variable. También denominada como serie cronológica.

$$Y_1, Y_2, Y_3, Y_4, \dots, Y_t$$

Se reconocen generalmente tres categorías de modelos de series temporales que se consideran adecuadas para aproximar las relaciones fundamentales entre los distintos componentes de los datos observados.

#### 2.2.5.1. Aditivo

Menciona Caridad y Ocerin (2005) que en los gráficos es usual ver una evolución, en el que la serie tiene tres componentes, los cuales son: Tendencia ( $T_t$ ), Estacional ( $E_t$ ), el o los ciclos ( $C_t$ ) y el componente irregular ( $I_t$ ), dando lugar al modelo aditivo.

$$Y_t = T_t + E_t + C_t + I_t$$

#### 2.2.5.2. Multiplicativo

De acuerdo con Hankhe y Reitsch (1996) en el caso de series temporales que se miden en períodos inferiores a un año, como las series mensuales o trimestrales, se asume que cada valor original resulta de la multiplicación de cuatro componentes distintos.

$$Y_t = T_t * E_t * C_t * I_t$$



### 2.2.5.3. Mixto

En series de tiempo, existen modelos clásicos mixtos, donde algunas de sus componentes operan de manera aditiva mientras que otras lo hacen de manera multiplicativa, se parecen a los multiplicativos. (Caridad y Ocerin, 2005)

$$Y_t = T_t * E_t * C_t + I_t$$

$$Y_t = (T_t + E_t + C_t) * I_t$$

Dónde:

- Tt: Componente de Tendencia.
- It: Componente Aleatorio o irregular
- Ct: Componente Cíclico.
- Yt: Serie observada en el instante t.
- Et: Componente Estacional.

### 2.2.6. Modelo

Un modelo es un instrumento que se sustenta en el análisis de datos para representar la realidad utilizando métodos estadísticos. Consiste en una representación matemática de los datos observados, con el objetivo de capturar con la mayor precisión posible la relación entre una o más variables, considerando tanto los datos empleados como la influencia que ejercen en las observaciones. (Contento Rubio, 2019)



### 2.2.7. Modelo Series de Tiempo

Según José (2009) los modelos de series de tiempo describen de manera adecuada la procedencia, los modelos de series temporales consisten en formas teóricas que pueden ser determinísticas, aleatorias o una combinación de ambas, utilizadas para analizar una serie de datos a lo largo del tiempo.

- Las variables temporales: Se entiende como aquellas que se observan y registran a lo largo de un período temporal. Por ejemplo,  $Y_t$  representa la variable  $Y$  en un momento específico  $t$ .
- Serie Temporal: Se trata de un conjunto de  $t$  observaciones, donde cada observación corresponde a una de las variables:  $Y_1, Y_2, \dots, Y_t$ . La serie cronológica, también conocida como serie temporal, hace referencia a las observaciones de la variable temporal, las cuales son denominadas "realizaciones".

### 2.2.8. Modelo Univariante

Menciona Jiménez et al., (2006) en el análisis univariante, no es necesario buscar relaciones causales que expliquen el comportamiento de la variable principal, ni obtener información sobre otras variables que puedan influir en ella. En este contexto, no se investiga ni se considera ninguna relación con otras variables explicativas.

La predicción univariante según Chatfield (1987) menciona este enfoque se emplea en cuestiones económicas, principalmente con dos propósitos principales:

- La proyección de ciertas variables explicativas de un modelo causal implica anticipar que en el futuro mantendrán ciertas características observadas en su evolución previa.
- La anticipación a corto plazo se destaca por su habilidad para capturar la dinámica del comportamiento de la variable analizada, especialmente en situaciones habituales. Esto implica que en situaciones donde no hay cambios drásticos con respecto a los datos previos de la variable, estas técnicas ofrecen pronósticos precisos y confiables.

Dentro de una serie temporal  $\{Y_t\}$ , los modelos univariantes se singularizan por tener una sola variable observada a lo largo del tiempo, expresándose comúnmente en forma polinomial. Dentro de este enfoque, se encuentran técnicas simples como el “modelo autorregresivo” de primer orden o modelos de tendencia lineal o exponencial. Sin embargo, las técnicas más sólidas para predecir una sola variable son aquellas asociadas con los modelos de Box-Jenkins, específicamente los modelos ARIMA. Estos últimos forman parte de un conjunto más amplio de técnicas desarrolladas por Box y Jenkins, que abarcan un espectro más diverso de métodos para el análisis y la predicción de series temporales. (Chinlli, 2021)

- Modelo Univariante de Box-Jenkins No Integrado
- Modelo Univariante de Box-Jenkins Integrado

### **2.2.9. Procesos Estocásticos**

Menciona Peña (2005) que la categorización de una serie como estacionaria o no estacionaria se encuentra intrínsecamente vinculada al lapso de tiempo bajo observación. Esto se debe a que una serie puede exhibir estabilidad en un periodo breve, pero volverse no estacionaria cuando se amplía el período de



análisis. Un ejemplo ilustrativo de este fenómeno es el seguimiento de la temperatura en una ubicación específica de la Tierra: durante un intervalo de 20 años, podría aparentar una estabilidad relativa, pero al extender el análisis a miles de años, podría manifestar una tendencia creciente. Como principio general, a medida que se prolonga el período de observación, aumenta la complejidad para que el proceso en cuestión se mantenga estable.

Al momento de observar una serie de tiempo, se asumirá que la serie es una realización de un proceso estocástico. El cual Matilla et al., (2017) define a un proceso estocástico como un conjunto de variables aleatorias en momentos sucesivos en el tiempo, es decir, dado un suceso determinado, que es observado cronológicamente, bajo ciertas condiciones de estabilidad temporal, los datos observados pueden permitir caracterizar al proceso generador de datos.

#### **2.2.9.1. Proceso Estocástico Estacionario**

González (2009) sostiene que, en el análisis de series temporales, el propósito consiste en aplicar la teoría de procesos estocásticos para identificar qué proceso produjo la serie temporal y entender cómo se comporta. Esto incluye proyectar su evolución futura. Sin embargo, para obtener predicciones confiables, no basta con elegir cualquier tipo de proceso estocástico; es fundamental que la estructura de probabilidad del proceso permanezca constante a lo largo del tiempo.

Según Caridad y Ocerin (2005) mencionan que un proceso estocástico se considera estacionario cuando, para todos los instantes de tiempo  $t$ , su media y varianza permanecen constantes, al igual que la autocorrelación a lo largo de la serie temporal.



Un proceso estocástico se vuelve estrictamente estacionario cuando sus características estadísticas o probabilísticas no experimentan variaciones en el tiempo, lo que implica que su distribución acumulativa no depende del tiempo (Montenegro, 2011).

La esencia de la teoría de la predicción radica en la búsqueda de regularidades en el comportamiento histórico de la serie temporal y su extrapolación hacia el futuro. Es crucial que los procesos estocásticos que generan las series de tiempo sean estables para que sean útiles en la predicción. Si muestran un comportamiento cambiante e inestable en cada momento, no son adecuados para la predicción. Estas condiciones, que requieren estabilidad en los procesos estocásticos, se conocen como estacionariedad (González, 2009).

Estacionariedad Estricta: Un proceso estocástico  $Y_t$ , es estacionario en sentido estricto si y solo si:

$$F[Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}] = F[Y_{t_1+k}, Y_{t_2+k}, \dots, Y_{t_n+k}] \quad \forall (t_1, t_2, \dots, t_n)$$

y  $k$

Esto significa que la distribución de un grupo de variables aleatorias del proceso permanece constante incluso cuando se desplazan  $k$  períodos en el tiempo.

Estacionariedad en Covarianza: Un proceso estocástico  $Y_t$ , es considerado estacionario en términos de covarianza si y solo si:



Esto significa que la distribución de un grupo de variables aleatorias del proceso permanece constante incluso cuando se desplazan  $k$  períodos en el tiempo:

$$E(Y_t) = \mu < \infty, \quad \forall t$$

Todas las variables aleatorias exhiben una varianza constante y finita, lo que evidencia que la dispersión alrededor de la media se mantiene continua a lo largo del tiempo para todas las variables del proceso.

$$V(Y_t) = E[Y_t - \mu]^2 = \sigma_Y^2 < \infty, \quad \forall t$$

Las autocovarianzas son independientes del tiempo; únicamente dependen del número de periodos entre las variables. Esto significa que la covarianza entre dos variables aleatorias que están separadas por  $k$  periodos es constante, sin importar el momento específico en el tiempo al que se refieran.

$$\text{cov}(Y_t Y_s) = E[Y_t - \mu][Y_s - \mu] = \gamma_{|t-s|} = \gamma_k < \infty, \quad \forall k$$

En consecuencia, un proceso estocástico exhibe estacionariedad en covarianza si:

$$E(Y_t) = \mu < \infty$$

$$\text{Cov}(Y_t Y_s) = \begin{cases} V(Y_t) = \sigma_Y^2 < \infty & t = s \\ \gamma_{|t-s|} = \gamma_k < \infty & t \neq s \end{cases}$$

Un proceso estocástico que es estacionario en covarianza se caracteriza por su conocimiento:

$$\mu, \quad V(Y_t) \text{ y } \gamma_k, k = \pm 1, \pm 2, \dots,$$

### 2.2.10. Ruido Blanco

El ruido blanco, considerado la serie más aleatoria, es esencial en el modelado de otras series temporales. En términos formales, se describe como un proceso estocástico en el que cada observación es independiente e idénticamente distribuida.

Según Matilla et al., (2017) el ruido blanco es un componente esencial en el análisis de series temporales y se considera uno de los procesos estocásticos estacionarios más simples y fundamentales. Se distingue por ser una secuencia de variables aleatorias con media 0, varianza constante e invariable, y covarianzas nulas entre ellas. Que responde a la siguiente expresión analítica:

$$Z_t = \bar{Z} + \varepsilon_t$$

Y puede ser escrito en desviaciones a las medias sin pérdida de generalidad, es decir que, expresado de esta forma, el proceso denominado Ruido Blanco sería:

$$z_t = \varepsilon_t$$

Y satisface:

$$1) E(z_t) = 0, \quad t = 1, 2, 3, \dots$$

$$2) \text{var}(z_t) = \sigma_\varepsilon^2, \quad t = 1, 2, 3, \dots$$

$$3) \text{cov}(z_t, z_{t+u}) = \gamma_u = 0 \text{ para todo } u \neq 0$$

La primera condición asegura que la media siempre permanece constante y es igual a cero. La segunda condición implica que la varianza permanece



invariable, mientras que la última establece que las variables del proceso están no correlacionadas en todos los desfases (Peña, 2005).

Montenegro (2011) nos da como ejemplo el lanzamiento de un dado. En este caso, la secuencia de lanzamientos constituye un ruido blanco, ya que cada lanzamiento es independiente de los anteriores y posteriores, y la probabilidad de obtener cualquier resultado en cada lanzamiento permanece constante. Sin embargo, si consideramos un dado cargado, aunque los lanzamientos aún puedan ser independientes e idénticamente distribuidos, la probabilidad ya no seguirá una distribución uniforme, ya que algunas caras del dado tendrán una probabilidad mayor de aparecer que otras. Esto ocurre especialmente si un jugador habilidoso puede influir en la forma en que el dado cae, la situación se vuelve menos clara. En este escenario, los lanzamientos pueden no ser repeticiones independientes del experimento, lo que hace menos probable que la serie sea considerada ruido blanco. Es importante destacar que el ruido blanco no es predecible basándose en su propio historial, ya que cada observación es independiente de las demás. Su predictibilidad es incondicional, lo que limita su utilidad en el proceso de pronóstico.

Por tanto, es una sucesión de variables aleatorias cuya media es cero y su varianza permanece constante. Por lo que Matilla García et al., (2017) concluye que el proceso de ruido blanco, conoce los valores pasados, pero no proporciona ninguna información sobre los valores futuros, el proceso no tiene memoria. Por lo cual, es evidente que el ruido blanco es un proceso estacionario débil.

### 2.2.11. Coeficiente de Correlación

La forma de medición comúnmente usada para evaluar el coeficiente de correlación es el nivel de asociación lineal entre dos variables específicas. Se encuentra dentro del rango de -1 a +1, indicando el grado de relación entre dos variables medidas. Cuando la correlación es -1, las variables tienen una correlación negativa perfecta. Por otro lado, una correlación positiva perfecta se representa con un coeficiente de +1. El coeficiente de correlación indica tanto la fuerza como la relación entre las variables, pudiendo ser desde una fuerte correlación negativa hasta una fuerte correlación positiva. Por lo que Hankhe y Reitsch (1996) concluye que el coeficiente de correlación evalúa el nivel de relación lineal que existe entre dos variables. Es una medida que indica cómo se comportan dos variables en conjunto en términos de su relación lineal.

Es crucial diferenciar entre los dos conjuntos de datos de interés para el pronosticador. En la población que abarca todos los puntos  $X - Y$  de interés, se utiliza un coeficiente de correlación representado por  $\rho$ . Por otro lado, cuando se toma una muestra de estos datos  $X - Y$ , el coeficiente de correlación de esta muestra se denomina como  $r$ . Hankhe y Reitsch (1996) a continuación se ofrece un compendio de los aspectos destacados de  $\rho$  y  $r$ .

**Tabla 1**

*Correlación de la Población y de la Muestra*

POBLACIÓN	MUESTRA
Coeficiente de correlación:	Coeficiente de
$\rho(-1 \leq r \leq 1)$	correlación: $r(-1 \leq r \leq 1)$

---

$\rho = -1$ : correlación negativa perfecta	$r = -1$ : correlación negativa perfecta
$\rho = 0$ : sin correlación	$r = 0$ : sin correlación
$\rho = 1$ : correlación positiva perfecta	$r = 1$ : correlación positiva perfecta

---

Nota: (Hankhe & Reitsch, 1996).

### 2.2.12. Autocorrelación

Cuando se observa una variable a lo largo del tiempo, es común que esté correlacionada consigo misma en desfases de uno o más periodos. El coeficiente de autocorrelación mide la relación entre una variable y ella misma, pero con un desplazamiento temporal. Este análisis se emplea para identificar patrones en los datos, como tendencias, estacionalidades o irregularidades. Al calcular los coeficientes de autocorrelación para distintos intervalos temporales, podemos detectar estos patrones en las series de tiempo (Hankhe y Reitsch, 1996).

#### 2.2.12.1. Función de Autocorrelación (FAC)

Según González (2009) refiere que un “coeficiente de autocorrelación de orden  $k$  en un proceso estocástico estacionario, lo que indica la fuerza de la relación lineal entre dos variables aleatorias del proceso, que están distanciadas por un intervalo de  $k$  periodos”.

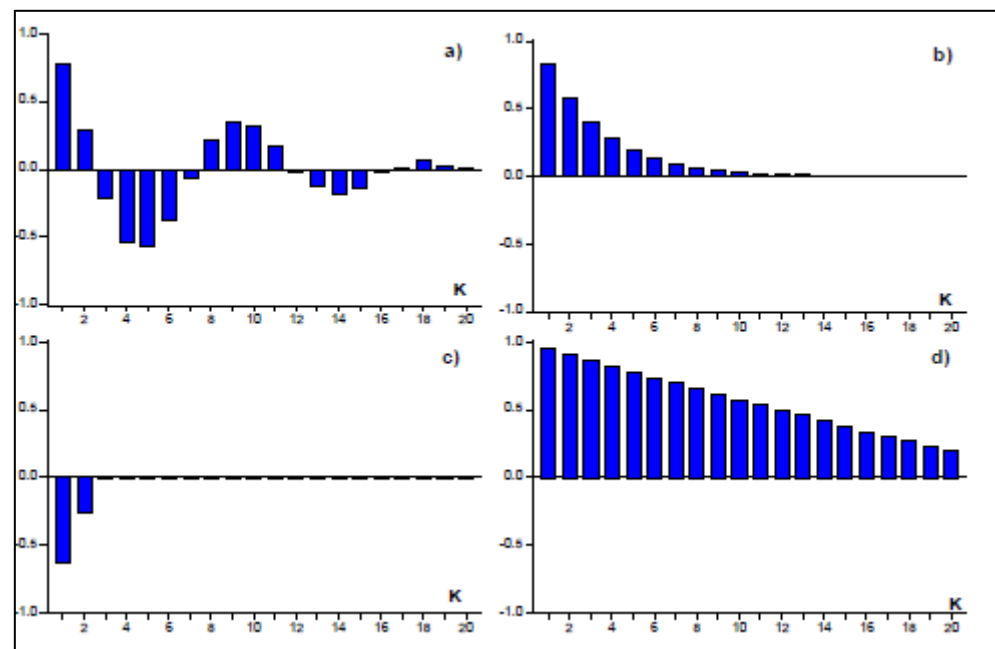
$$\rho_k = \frac{cov(Y_t Y_{t+k})}{\sqrt{V(Y_t)V(Y_{t+k})}} = \frac{\gamma_k}{\sqrt{\gamma_0\gamma_0}} = \frac{\gamma_k}{\gamma_0}$$

Al ser un coeficiente de correlación, su valor no está influenciado por las unidades utilizadas  $|\rho_k| \leq 1, \forall k$ .

La función de autocorrelación de un proceso estocástico estacionario, que se define como una función que varía según el intervalo  $k$ , donde  $k$  representa los diferentes desfases en el tiempo. Esta función, denotada como  $\rho_k$ , con  $k = 0,1,2,3,\dots$ . El correlograma, comúnmente utilizado, representa todos los coeficientes de autocorrelación del proceso, proporcionando una visualización de la función de autocorrelación.

**Figura 2**

*Funciones de Autocorrelación*



Fuente: (González, 2009).

Una función de autocorrelación de un proceso estocástico estacionario presenta algunas propiedades particulares:

Por definición, el coeficiente de autocorrelación de orden cero siempre es igual a 1. Por consiguiente, suele omitirse en la función de autocorrelación, ya que es constante y no proporciona información adicional sobre la autocorrelación en diferentes desfases.



Es una función simétrica:  $\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_{-k}}{\gamma_0} = \rho_{-k}$ . Por esta razón, en el correlograma se visualiza la función de autocorrelación únicamente para los retrasos positivos  $k$ .

En los procesos estocásticos estacionarios, la función de autocorrelación tiende a cero rápidamente a medida que  $k$  aumenta hacia  $\infty$ .

La función de autocorrelación es crucial para capturar los gráficos en la figura representan la estructura dinámica lineal del modelo. Los correlogramas de la figura señalan cuatro series temporales distintas a), b), c) y d) disminuyen rápidamente hacia cero a medida que  $k$  aumenta: de manera exponencial En los casos a) y b), y cortándose en el caso c). Estos correlogramas representan series estacionarias. En contraste, los coeficientes de autocorrelación en el correlograma d) disminuyen gradualmente y de forma constante, lo que sugiere que no se trata de una serie estacionaria (González, 2009).

Para describir un proceso estocástico estacionario por sus momentos, se pueden considerar dos alternativas:

La media y la función de autocovarianza:  $\mu, \gamma_k, k = 0, 1, 2, 3, \dots$

Media, dispersión y función de autocorrelación:  
 $\mu, \gamma_0, \rho_k, k = 0, 1, 2, 3, \dots$



### 2.2.12.2. Función de Autocorrelación Parcial (FACP)

Peña (2005) refiere que una función de autocorrelación parcial de un proceso estacional multiplicativo se complica debido a que involucra una combinación de las funciones de autocorrelación regular y estacional, también se considera la interacción en los retrasos estacionales, lo que revela la estructura de la sección regular. La función de autocorrelación parcial se visualiza con la función de autocorrelación simple, pareciéndose a esta última. En el caso específico de un proceso estacional, la función de autocorrelación parcial sigue este mismo patrón:

La función de autocorrelación parcial (FACP) reflejará la estructura regular en los primeros retrasos, mientras que los retrasos estacionales mostrarán la FACP estacional.

Cada coeficiente estacional se encuentra acompañado por una autocorrelación parcial a su derecha. (retardos  $js + 1, js + 2, \dots$ ) en el caso de coeficientes estacionales positivos, la estructura regular de la FACP se presenta invertida, mientras que, para coeficientes negativos, la FACP aparece con su respectivo signo.

En el lado izquierdo de cada coeficiente estacional (retardos  $js - 1, js - 2, \dots$ ), Podemos examinar la función de autocorrelación simple de la componente regular. En los modelos estacionales, los retrasos estacionales son considerablemente menores en magnitud que los que se verían en una estructura estacional pura.



De acuerdo con (Matilla García et al., 2017) la función de autocorrelación ayuda a caracterizar a los procesos  $AR$ : en general, si decrece rápidamente nos encontraremos, ante un proceso  $AR$ . Sin embargo, la función de autocorrelación no nos informa del *orden del proceso autorregresivo*. Para determinarlo, es decir para identificar el orden  $p$  de un proceso  $AR(p)$  debemos recurrir a la función de autocorrelación parcial (FACP).

La autocorrelación parcial con  $k$  desfases, el cual será denominado como  $\phi_{kk}$ , mide la influencia de  $z_{t-k}$  sobre  $z_t$  descontada la influencia de los  $k - 1$  valores anteriores de  $z$ :  $z_{t-1}, z_{t-2}, \dots, z_{t-k-1}$ . Es decir, la autocorrelación parcial de  $k$  se refiere a la correlación entre  $z_t, z_{t-k}$  condicionada a  $z_{t-1}, z_{t-2}, \dots, z_{t-k-1}$ .

Esta correlación será dada por un coeficiente  $\phi_{kk}$ , que irá definiendo la función de autocorrelación parcial para los desfases temporales 1 a  $k$  de cualquier proceso estacionario.

### 2.2.13. Evaluación de Modelos $AR$ , $MA$ , $ARMA$ y $ARIMA$

Menciona González (2009) cuando se establece un modelo univariante de series de tiempo, no necesariamente el objetivo prima en obtener el modelo "verdadero". Se trata más bien de lidiar con la complejidad de modelar una realidad multifacética. El objetivo principal es desarrollar un modelo que sea simple, pero al mismo tiempo preciso en la representación de las características clave de la serie, especialmente su estructura de autocorrelación. Por otra parte, los modelos  $ARMA(p,q)$ ;  $AR(p)$  y  $MA(q)$  son simplificaciones del modelo lineal

general y se utilizan para entender las propiedades de la autocorrelación de la serie. Posteriormente, se aplican en la modelización y predicción de las series.

Según Montenegro (2011) es frecuente encontrar que, en muchos fenómenos temporales, el pasado ofrece señales sobre lo que podemos esperar en el futuro. El historial de una serie temporal a menudo contiene y reemplaza la información de otras variables que podrían influir en el proceso. Una técnica de modelado bastante simple pero efectiva, que sustenta en expresar una variable en función de su propio historial, es conocida como autorregresión. Esta técnica forma parte del enfoque de modelado propuesto por Box-Jenkins. Dentro de este marco, se destacan dos formulaciones principales: los “modelos autorregresivos”  $AR(p)$  y los modelos de promedio móvil  $MA(q)$ , los cuales están diseñados específicamente para procesos estocásticos univariados estacionarios.

#### 2.2.14. Modelo Lineal Estacionario

##### 2.2.14.1. Modelo Autorregresivo (AR)

###### Modelo Autorregresivo $AR(p)$

El “modelo autorregresivo” representa el valor presente de una serie estacionaria  $x_t$  basándose en su propia historia, es decir, en sus rezagos  $x_{t-1}, \dots, x_{t-p}$ . Un “modelo autorregresivo”  $AR(p)$  se expresa algebraicamente de la siguiente manera:

$$x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} = \varepsilon_t + c$$

En esta expresión,  $c$  representa un valor constante, mientras que  $p$  indica el orden o la máxima cantidad de períodos anteriores considerados en el modelo, y  $\varepsilon_t$  Se refiere a una serie de valores aleatorios que exhiben

propiedades específicas: su media es cero y su varianza permanece constante en el tiempo  $\sigma_{\varepsilon}^2$ . En los estudios sobre series temporales, esta sucesión de datos se describe comúnmente con diversos términos, como ruido, error, residuos, innovaciones o shocks  $\{\varepsilon_t\}$  en la literatura especializada en análisis de series temporales, este conjunto de datos suele denominarse de manera intercambiable como ruido, errores, residuos, innovaciones o shocks. Otra presentación usual es con  $x_t$  despejada. (Montenegro, 2011)

Según González (2009) sostiene que un “modelo autorregresivo” finito de orden  $p$ , AR ( $p$ ) es una aproximación lógica y comúnmente utilizada al modelo lineal general, expresa  $Y_t$  basándonos en su historial hasta el momento del retardo  $t - p$ . Obtenemos un modelo finito simplemente recortando el modelo general:

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

Modelo autorregresivo de primer orden AR (1)

De acuerdo con Montenegro (2011) es uno de los “modelos autorregresivos” más sencillos, el de orden 1, que tiene solo un rezago y se escribe de la siguiente manera:

$$x_t = c + \alpha x_{t-1} + \varepsilon_t$$

Usando el operador de retraso, como  $(1 - \alpha L)x_t = c + \varepsilon_t$ .

La media del modelo AR(1) se puede calcular mediante la toma de expectativas de la formula dada anteriormente y recordando que, por el supuesto de estacionalidad,  $Ex_t = Ex_{t-1}$ ,

$$Ex_t = E(c + \alpha x_{t-1} + \varepsilon_t) = c + \alpha Ex_{t-1} + E\varepsilon_t$$

De donde, reorganizando, se tiene:

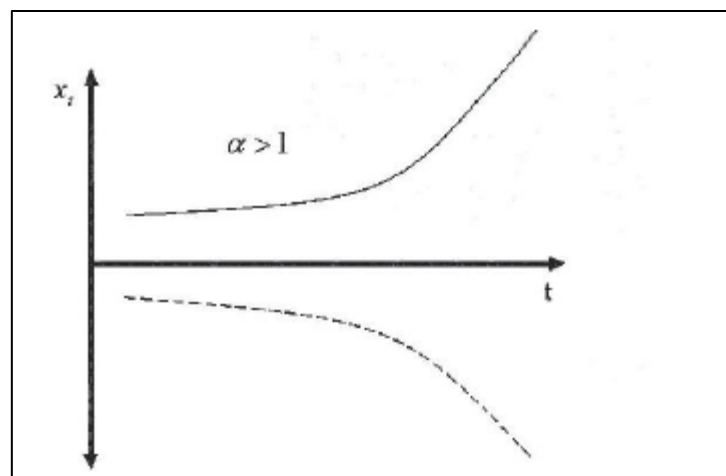
$$Ex_t = \frac{c + E\varepsilon_t}{1 - \alpha} = \frac{c}{1 - \alpha}$$

Se supuso que  $E\varepsilon_t = 0$ .

Es importante observar que en la primera fórmula es necesario que  $|\alpha| < 1$  para que  $x_t$  es necesario recordar Podemos expresar que  $\alpha$  representa el inverso de la raíz de la ecuación auxiliar  $(1 - \alpha L) = 0$  y para asegurar la estabilidad del sistema, esta raíz debe tener un valor absoluto mayor que uno. De lo contrario, que  $|\alpha| > 1$  entonces la serie crece sin límite como se muestra en la figura.

### Figura 3

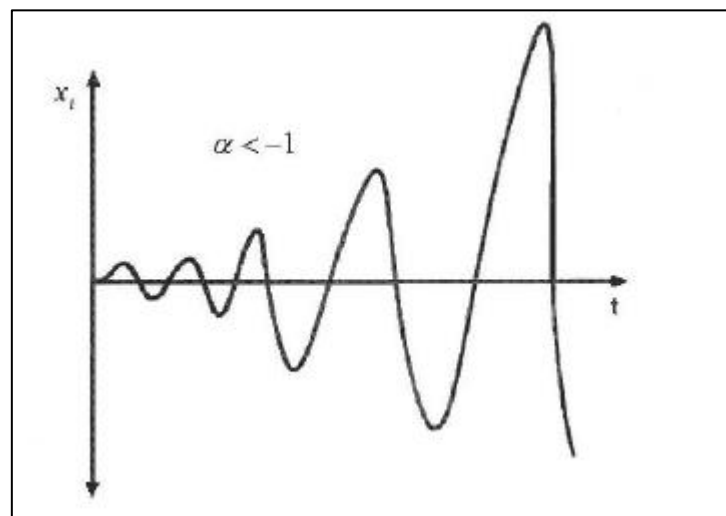
*AR(1) con  $\alpha > 1$*



Nota: (González Casimiro, 2009).

**Figura 4**

*AR(1) con  $\alpha < -1$*



Nota: (González Casimiro, 2009).

La figura muestra, con  $\alpha > 1$  dependiendo del valor de  $x_0$  y de los  $\varepsilon_t$  iniciales,  $x_t$  seguirá la curva sólida o la línea punteada. Y en la siguiente figura, el coeficiente  $\alpha < -1$  induce un comportamiento oscilatorio creciente en la serie.

Según (González Casimiro, 2009) el proceso AR(p) se diferencia por satisfacer las condiciones de estacionariedad para cualquier conjunto de parámetros. Además, un proceso autorregresivo finito AR(p) es considerado estacionario si sus raíces del polinomio autorregresivo  $\phi_p(L)$  tienen magnitud mayor que uno, lo que implica que no están dentro del círculo unidad.

Condiciones de estacionariedad:

Para el Modelo AR(1):

$$Y_t = \phi Y_{t-1} + \alpha_t \rightarrow (1 - \phi L) Y_t = \alpha_t$$

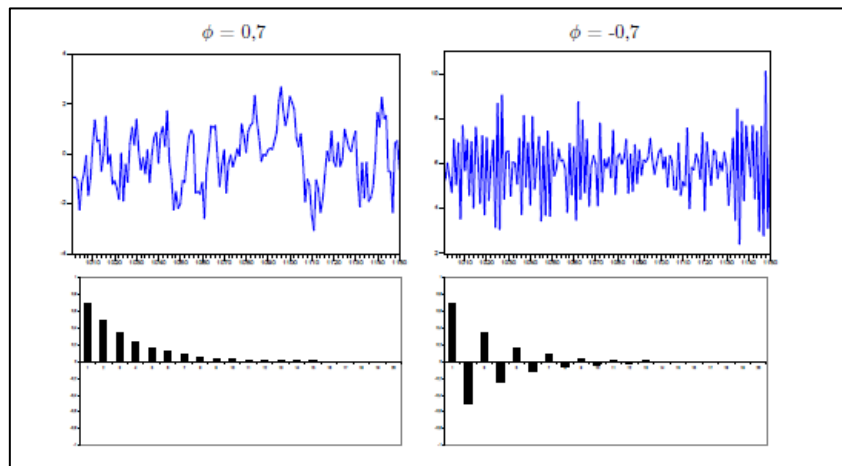
Polinomio autorregresivo:  $\phi_1(L) = 1 - \phi L$

Raíces:  $1 - \phi L = 0 \rightarrow L = \frac{1}{\phi}$

La condición de estacionariedad:  $|L| = \left| \frac{1}{\phi} \right| > 1 \rightarrow |\phi| < 1$

### Figura 5

*Procesos autorregresivos de orden 1*



Nota: (González Casimiro, 2009).

Para el Modelo AR(2)

$$Y_t = \phi Y_{t-1} + \phi Y_{t-2} + \alpha_t \rightarrow (1 - \phi_1 L - \phi_2 L^2) Y_t = \alpha_t$$

El polinomio de autorregresión:  $\phi_2(L) = 1 - \phi_1 L - \phi_2 L^2$

Raíces:  $1 - \phi_1 L - \phi_2 L^2 = 0 \rightarrow L_1, L_2 = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4 * \phi_2}}{-2 * \phi_2}$

Condición de estacionariedad:

$$|L_1| = \left| \frac{\phi_1 + \sqrt{\phi_1^2 + 4 * \phi_2}}{-2 * \phi_2} \right| > 1 \rightarrow |\phi| < 1$$

$$|L_2| = \left| \frac{\phi_1 - \sqrt{\phi_1^2 + 4 * \phi_2}}{-2 * \phi_2} \right| > 1 \rightarrow |\phi| < 1$$

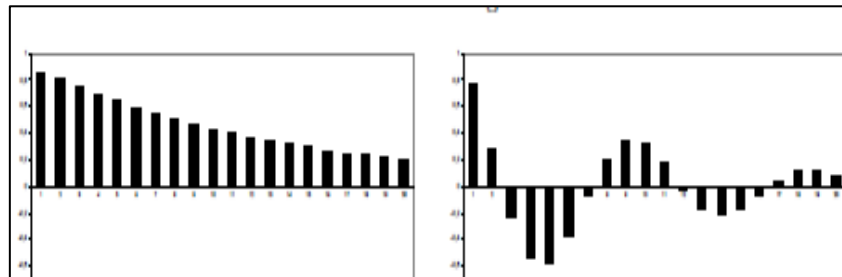
¿Cómo se calcula el módulo de estas raíces?

Si el radicando es positivo,  $(\phi_1^2 + 4 * \phi_2) > 0$ , las raíces son reales y el modulo es el valor absoluto.

Si el radicando es negativo,  $(\phi_1^2 + 4 * \phi_2) < 0$ , las raíces son complejas de la forma  $L_1, L_2 = \alpha \pm bi$  y su modulo es igual a :  $|L_1| = |L_2| = \sqrt{\alpha^2 + b^2}$

### Figura 6

*Procesos autorregresivos de orden 2*



#### 2.2.14.2. Modelo de Medias Móviles (MA)

Montenegro (2011) entró de la “metodología de Box-Jenkins” se incluye otro modelo, conocido como promedio móvil. En este modelo, la serie  $x_t$  se expresa como una combinación del presente y el pasado de una serie de ruido blanco  $\{\varepsilon_t\}$  la cual tiene una media de cero y una varianza finita:

$$x_t = c + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} = c + \sum_{s=0}^q \beta_s \varepsilon_{t-s}$$

Según (Montenegro García, 2011) un modelo de medias móviles de orden finito  $q$ , denominado MA( $q$ ), constituye una simplificación del modelo lineal general al truncar el modelo de medias móviles de orden infinito:

$$Y_t = \alpha_t - \theta_1\alpha_{t-1} - \theta_2\alpha_{t-2} - \dots - \theta_q\alpha_{t-q} \quad \alpha_t \sim RB(0, \sigma^2)$$

#### Medias Móviles MA( $q$ )

Este modelo describe el valor de  $Y_t$  en relación con la innovación actual y sus valores pasados hasta el retardo  $q$ :

$$Y_t = \alpha_t - \theta_1\alpha_{t-1} - \theta_2\alpha_{t-2} - \dots - \theta_q\alpha_{t-q} \quad \alpha_t \sim RB(0, \sigma^2)$$

En términos del operador de retraso, la expresión es la siguiente:

$$Y_t = (1 - \theta_1L - \theta_2L^2 - \theta_3L^3 - \dots - \theta_qL^q)\alpha_t \quad Y_t = \theta_q(L)\alpha_t$$

En esta ecuación, el término  $\theta_q(L)$  representa el polinomio de medias móviles, y el vector  $(\theta_1, \theta_2, \theta_3, \dots, \theta_q)$  corresponde a los parámetros de las medias móviles.

Un modelo MA( $q$ ) se extiende del modelo MA(1), lo que quiere decir que comparten similitudes en sus características. No obstante, al incorporar más retardos de la perturbación al modelo, su capacidad de retención de información aumenta y la complejidad de la estructura dinámica que describe es mayor. Si una perturbación ocurre en el tiempo  $t$ , esta afecta a  $Y_t$ , en  $Y_{t+1}$  y a los valores subsiguientes hasta  $Y_{t+q}$ .

Entonces, la perturbación  $\alpha_t$  en un modelo MA( $q$ ) exhibe una memoria prolongada en el sistema durante  $q$  periodos. Esta persistencia



prolongada del MA(q) se reflejará en la configuración de la función de auto covarianzas y/o auto correlación.

El modelo MA(q) exhibe una media constante y nula, una varianza constante y finita, y su estructura de autocovarianza se limita a partir del retardo q, lo que significa:

$$\rho_k = \begin{cases} \rho_k \neq 0 & k = 1, 2, \dots, q \\ \rho_k = 0 & k > q \end{cases}$$

El primer modelo mostrado puede ser extendido de manera sencilla para describir modelos que no tienen media nula:

$$Y_t = \delta + \alpha_t - \theta_1 \alpha_{t-1} - \theta_2 \alpha_{t-2} - \dots - \theta_q \alpha_{t-q} \quad \alpha_t \sim RB(0, \sigma^2)$$

#### Medias Móviles MA(1)

Este modelo MA(1) define el valor de Y en el instante t basándose en la innovación actual y su retardo inmediato:

$$Y_t = \alpha_t - \theta_1 \alpha_{t-1} \quad \alpha_t \sim RB(0, \sigma^2) \quad t = 1, 2, \dots, \theta$$

A menudo, se refiere a los procesos de medias móviles a los modelos MA se les conoce por tener una memoria de corto plazo, en contraste, a los autorregresivos se les caracteriza como procesos de memoria prolongada.

Esto se debe a la estructura del modelo MA(1) donde la perturbación  $\alpha_t$  se hace evidente en el sistema en el instante t y afecta solo a  $Y_t$  e  $Y_{t+1}$  lo que significa que su influencia se limita a un solo periodo de tiempo. Por otro lado, en un proceso autorregresivo, la perturbación  $\alpha_t$  se

presenta en el sistema en el momento  $t$  y afecta a  $Y_t$  y a través de  $Y_t$  influye en las observaciones futuras, lo que implica que su efecto en el sistema se prolonga durante varios periodos de tiempo.

$$\begin{array}{cccccccc} \dots & Y_{t-2} & Y_{t-1} & Y_t & Y_{t+1} & Y_{t+2} & \dots \\ \nearrow & \uparrow & \nearrow & \uparrow & \nearrow & \uparrow & \nearrow \\ \dots & \alpha_{t-2} & \alpha_{t-1} & \alpha_t & \alpha_{t+1} & \alpha_{t+2} & \dots \end{array}$$

Vamos a verificar si el modelo MA(1) satisface los requisitos de estacionariedad para todos los valores del parámetro  $\theta$ .

Estacionario en media:

$$E(Y_t) = E(\alpha_t - \theta\alpha_{t-1}) = 0$$

En consecuencia, el modelo MA(1) se considera estacionario en términos de media para cualquier valor que pueda tomar el parámetro  $\theta$ .

Estacionario en covarianza:

$$\begin{aligned} \gamma_0 &= E(Y_t - E(Y_t))^2 = E(Y_t)^2 = E(\alpha_t - \theta\alpha_{t-1})^2 = \\ &= E(\alpha_t)^2 + \theta^2 E(\alpha_{t-1})^2 - 2\theta E(\alpha_t\alpha_{t-1}) \\ &= \sigma^2 + \theta^2\sigma^2 - 0 \end{aligned}$$

$$\gamma_0 = (1 + \theta^2)\sigma^2 < \infty$$

Es decir, la variabilidad del modelo MA(1) permanece constante y limitada para cualquier valor que pueda adoptar el parámetro  $\theta$ . Esto implica que la dispersión de los datos se mantiene estable independientemente de las diferentes configuraciones del parámetro  $\theta$ .

Las autocovarianzas,  $\gamma_k, k > 0$ :

$$\begin{aligned}\gamma_1 &= E(Y_t - E(Y_t))(Y_{t-1} - E(Y_{t-1})) = E(\alpha_t - \theta\alpha_{t-1})(\alpha_{t-1} - \theta\alpha_{t-2}) \\ &= \\ &= E(\alpha_t\alpha_{t-1}) - \theta E(\alpha_{t-1})^2 - \theta E(\alpha_t\alpha_{t-2}) + \theta^2 E(\alpha_{t-1}\alpha_{t-2}) \\ &= -\theta\sigma^2 < \infty\end{aligned}$$

$$\begin{aligned}\gamma_2 &= E(Y_t - E(Y_t))(Y_{t-2} - E(Y_{t-2})) = E(\alpha_t - \theta\alpha_{t-1})(\alpha_{t-2} - \theta\alpha_{t-3}) \\ &= \\ &= E(\alpha_t\alpha_{t-2}) - \theta E(\alpha_{t-1}\alpha_{t-2}) - \theta E(\alpha_t\alpha_{t-3}) + \theta^2 E(\alpha_{t-1}\alpha_{t-3}) = 0\end{aligned}$$

La función de autocovarianza de un proceso de medias móviles de orden  $MA(1)$  es:

$$\gamma_k = \begin{cases} \gamma_0 = (1 + \theta^2)\sigma^2 & k = 0 \\ \gamma_1 = -\theta\sigma^2 & k = 1 \\ \gamma_k = 0 & k > 1 \end{cases}$$

La función de autocovarianzas es finita y solo depende de la distancia  $k$ , no del tiempo, sin importar el valor que pueda tomar el parámetro  $\theta$ . De esta manera, podemos concluir que no es preciso imponer limitaciones al el valor del parámetro  $\theta$  debe ser tal que asegure la estacionariedad del modelo  $MA(1)$ . El modelo es robusto y puede ser aplicado de manera efectiva en una variedad de contextos sin necesidad de ajustes adicionales.

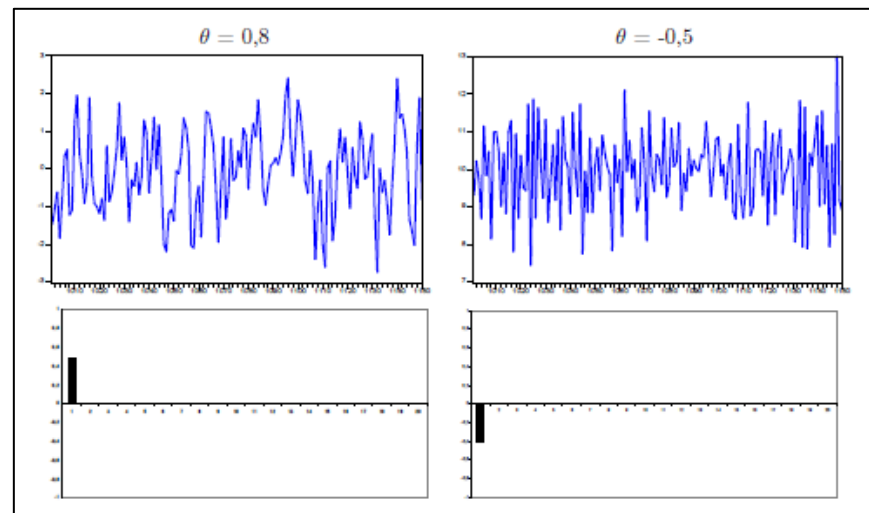
La ecuación de un modelo de medias móviles de orden  $MA(1)$  es:

$$\rho_k = \begin{cases} 1 & k = 0 \\ -\theta & k = 1 \\ \frac{1 + \theta^2}{0} & k > 1 \end{cases}$$

Una función de autocorrelación MA(1) se define por ser una función que se corta después de un cierto número de retardos experimenta una caída abrupta, alcanzando un valor de cero a partir de un solo rezago. Esto indica que las observaciones en un proceso MA(1) están débilmente correlacionadas con las observaciones pasadas, excepto con el valor inmediatamente anterior.(Chinlli, 2021)

### Figura 7

*Procesos Medias Móviles de orden 1:MA(1)*



En el gráfico se presentan dos instancias de los procesos MA(1) junto con sus respectivas funciones de autocorrelación asociadas. Este gráfico ofrece una representación visual de cómo se comportan estos procesos y cómo varían sus autocorrelaciones a través de diferentes intervalos de tiempo:

$$Y_t = \alpha_t + 0.8\alpha_{t-1}$$

$$Y_t = \alpha_t - 0.5\alpha_{t-1}$$

Ambas funciones de autocorrelación exhiben un corte en el primer rezago, una característica distintiva de un proceso de medias móviles de orden 1 es que su coeficiente de autocorrelación,  $\rho_1$ , puede ser negativo o positivo, dependiendo del signo del parámetro de medias móviles. Al examinar las series generadas por estos modelos, se nota que ambas fluctúan alrededor de la media cero con una variabilidad constante. La serie resultante del primer modelo, caracterizado por un parámetro positivo, parece exhibir una suavidad perceptible en contraste con la serie derivada del segundo modelo, identificado por un parámetro negativo. Este análisis nos proporciona una comprensión más profunda de cómo se comportan estos procesos y cómo influyen los parámetros en las características de las series temporales resultantes.

#### **2.2.14.3. Modelo Autorregresivo de Medias Móviles (ARMA)**

Según González (2009) los “modelos autorregresivos” de medias móviles establecen el valor de  $Y_t$  basándose en su historial hasta el rezago  $p$ , la innovación actual y el historial de la innovación hasta el rezago  $q$ . Esto implica que la variable  $Y_t$  es influenciada por su propio comportamiento pasado, así como por las fluctuaciones presentes y pasadas en la innovación del proceso. Este enfoque permite capturar tanto la dependencia temporal de la serie como la influencia de los factores externos en su evolución a lo largo del tiempo:

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + a_t + \phi_1 a_{t-1} + \dots \\ + \phi_q a_{t-q} \quad a_t \sim RB(0, \sigma^2)$$

La expresión de este modelo puede ser representada utilizando el operador de retardos de la siguiente manera:

$$(1 - \phi_1 L - \dots - \phi_p L^p) Y_t = (1 - \phi_1 L - \dots - \phi_q L^q) a_t$$

$$\phi_p(L) Y_t = \phi_q(L) a_t$$

Donde  $\phi_p(L)$  denota el polinomio autorregresivo, mientras que  $\phi_q(L)$  representa el polinomio de medias móviles.

El primer modelo representa una aproximación finita al modelo lineal general, tanto en su forma  $AR(\infty)$  como  $MA(\infty)$ . En realidad, cuando el modelo es estacionario, su representación  $MA(\infty)$  es la siguiente:

$$\begin{aligned} Y_t &= \frac{\theta_q(L)}{\phi_p(L)} a_t \rightarrow Y_t = \psi_{\infty}(L) a_t \rightarrow Y_t \\ &= a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \psi_3 a_{t-3} + \dots \end{aligned}$$

Tanto los coeficientes de la representación infinita de medias móviles (MA) como los de la forma infinita de autorregresión (AR) están limitados a depender del conjunto finito de parámetros ARMA. Esto significa que la estructura de los procesos ARMA influye en la distribución de los coeficientes en ambas representaciones, proporcionando un marco general para describir y comprender la dinámica de los “modelos autorregresivos” y de medias móviles  $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ .

Estacionariedad de un proceso ARMA (p, q):

El primer teorema establece que un proceso ARMA (p, q) es estacionario si y solo si las raíces del polinomio autorregresivo están fuera del círculo unitario. Esto significa que las raíces del polinomio  $\phi_p(L)$  (la



parte autorregresiva) deben estar fuera del círculo unitario en el plano complejo para que el proceso sea estacionario.

Ya que la parte de medias móviles finita siempre es estacionaria. Es decir, la estacionariedad del modelo ARMA está determinada principalmente por su parte autorregresiva. (Bernardí Cabrer, 2004)

Verificar la estacionariedad implica analizar las raíces del polinomio autorregresivo, lo que proporciona un método sistemático para evaluar esta propiedad del proceso ARMA.

Invertibilidad de un proceso ARMA (p, q):

El segundo teorema establece que un proceso ARMA (p, q) es invertible si y solo si las raíces del polinomio de medias móviles están fuera del círculo unitario. Esto implica que las raíces del polinomio de medias móviles  $\phi_p(L)$  deben estar fuera del círculo unitario en el plano complejo para que el proceso sea invertible.

A diferencia de la estacionariedad, la condición de invertibilidad se enfoca en la parte de medias móviles del modelo ARMA, ya que la parte autorregresiva finita siempre es invertible al estar expresada directamente en forma autorregresiva. (Guerra Rodríguez, 2022)

La verificación de la invertibilidad implica analizar las raíces del polinomio de medias móviles, lo que proporciona un método para evaluar esta propiedad del proceso ARMA.

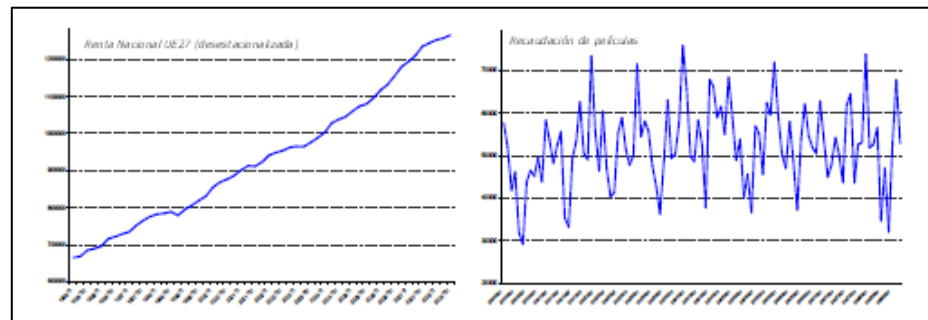
Estos teoremas proporcionan criterios fundamentales para determinar la estacionariedad y la invertibilidad de un proceso ARMA (p,

q), lo que contribuye a comprender mejor su comportamiento dinámico y sus propiedades estadísticas.

Los modelos mencionados anteriormente están fundamentados en la premisa de estacionariedad en covarianza, lo que implica que los procesos muestran una media y una varianza constantes y finitas, y las autocovarianzas permanecen constantes con el tiempo, cambiando solo en función del número de períodos de separación entre las variables (González, 2009).

### Figura 8

*Series temporales económicas*



Modelo autorregresivo de promedio móvil integrado ARIMA(p, d, q)

Supongamos el siguiente modelo ARMA(p,q):

$$\Phi_p(L)Y_t = \Theta_q(L)\alpha_t$$

El polinomio autorregresivo puede ser descompuesto en términos de sus p raíces  $L_1, L_2, \dots, L_p$

$$\Phi_p(L) = (1 - L_1^{-1}L)(1 - L_2^{-1}L) \dots (1 - L_p^{-1}L)$$



Imaginemos que  $(p-1)$  raíces son estacionarias, es decir, tienen un módulo fuera del círculo unitario, y una raíz es unitaria,  $L_i=1$ . En este caso, el polinomio autorregresivo se puede expresar de la siguiente manera:

$$\Phi_p(L) = (1 - L_1^{-1}L)(1 - L_2^{-1}L) \dots (1 - L_p^{-1}L) = \varphi_{p-1}(L)(1 - (1)^{-1}L)$$

$$\Phi_p(L) = \varphi_{p-1}(L)(1 - L)$$

### Estrategia de modelización ARIMA

Si los parámetros del modelo teórico  $\Phi(L)(1 - L)^d Y_t = \delta + \Theta(L)\alpha_t$  para el proceso  $Y_t$ , están disponibles, entonces, a partir de una secuencia concreta del ruido blanco,  $\alpha_1, \alpha_2, \dots, \alpha_T$ , y con valores iniciales para  $Y$  y  $\alpha$  se puede generar la serie temporal  $Y_1, Y_2, \dots, Y_T$ , q la cual representa una realización de tamaño T del proceso estocástico.

La estructura es, por lo tanto:

$$\text{Proceso estocástico} \begin{cases} \Phi(L)(1 - L)^d Y_t = \delta + \Theta(L)\alpha_t \\ \alpha_t \sim \text{RBN}(0, \sigma^2) \end{cases}$$

↓

Generación

↓

Realización:  $Y_1, Y_2, \dots, Y_T$

A partir de una misma configuración ARIMA(p,d,q), se pueden generar infinitas series. Cada serie estará acompañada por una secuencia distinta de ruido blanco,  $\alpha_1, \alpha_2, \dots, \alpha_T$ , y por ende, la serie temporal resultante,  $Y_1, Y_2, \dots, Y_T$ , variará en cada caso. No obstante, todas las series

temporales generadas bajo una misma configuración compartirán una similitud en su comportamiento dinámico.

En contraste, en la metodología Box-Jenkins, el proceso comienza de manera opuesta: se dispone de los valores de la serie temporal  $Y_1, Y_2, \dots, Y_T$  y se busca determinar la estructura ARIMA(p,d,q) que pudo haberla originado.

*Realización:*  $Y_1, Y_2, \dots, Y_T$

↓

*Inferencia*

↓

$$\text{Proceso estocástico} \begin{cases} \Phi(L)(1-L)^d Y_t = \delta + \Theta(L)\alpha_t \\ \alpha_t \sim \text{RBN}(0, \sigma^2) \end{cases}$$

La elaboración de modelos ARIMA sigue un proceso iterativo que comprende cuatro fases distintivas.

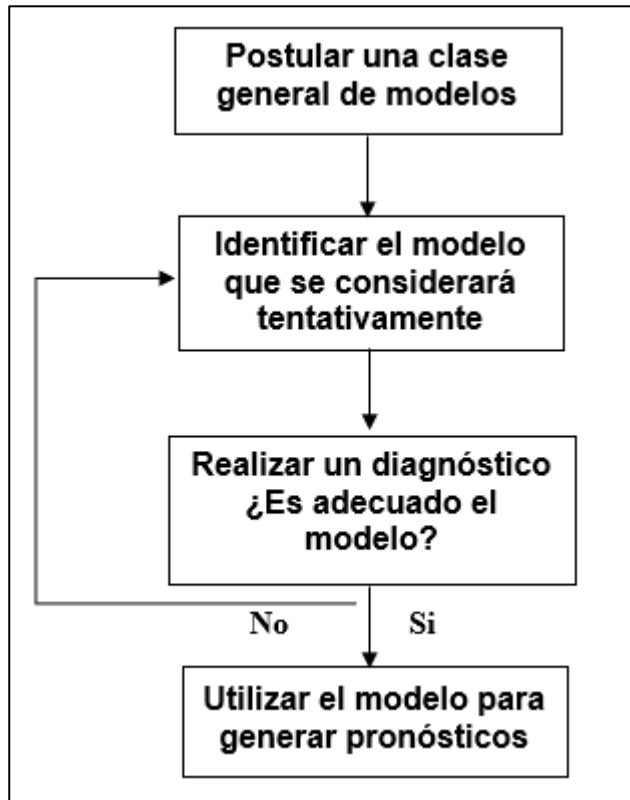
### 2.2.15. Metodología Box – Jenkins

Según Hankhe y Reitsch (1996) la “metodología de Box-Jenkins” abarca “una serie de pasos destinados a identificar, ajustar y validar modelos ARIMA utilizando datos de series temporales y que no sigue un patrón específico para prever los datos históricos de las series temporales; las predicciones provienen directamente de la forma del modelo ajustado. Se emplea un enfoque iterativo para establecer un modelo potencial dentro de una orden general de modelos. Una vez elegido un modelo, se compara con los datos cruciales para verificar si detalla con precisión la serie”. Si los residuos son pequeños, se distribuyen de manera aleatoria y no contienen información útil, lo que indica un ajuste adecuado del modelo. Sin embargo, si el modelo no se ajusta bien a la serie, el proceso se repite

utilizando un nuevo modelo, y así sucesivamente, hasta encontrar el modelo más apropiado.

**Figura 9**

*Metodología Box – Jenkins*



La “metodología de Box-Jenkins” se compone de cuatro fases para su realización del modelo, los mismos que son:

Paso 1: Identificación del modelo

Este paso permite determinar si la serie es estacionaria, lo que significa que varía alrededor de un nivel constante, o si es no estacionaria, indicando que la serie parece tener una tendencia creciente o decreciente a lo largo del tiempo y las autocorrelaciones no disminuyen rápidamente. (Santiago de la Fuente, 2012)

Es crucial que, si nuestra serie resulta ser no estacionaria, debemos aplicar pruebas de estacionariedad sustituyendo la serie original por una serie de diferencias. Esto implica modelar los cambios en lugar de los niveles. Por ejemplo, si la serie original muestra una tendencia de crecimiento con el tiempo, las primeras diferencias  $\Delta Y_t = Y_t - Y_{t-1}$ , Sería apropiado representar las variaciones a un nivel constante mediante un modelo ARMA. En ciertas ocasiones, es necesario aplicar una doble diferenciación para obtener datos estacionarios. Cuando se realiza una diferenciación dos veces, los datos estacionarios resultantes son:

$$\Delta^2 Y_t = \Delta(\Delta Y_t) = \Delta(Y_t - Y_{t-1}) = Y_t - 2Y_{t-1} + Y_{t-2}$$

El procedimiento continúa hasta que la representación gráfica de los datos muestre que la serie fluctúa alrededor de un nivel constante y las autocorrelaciones de la muestra desaparezcan rápidamente. Una vez que se obtiene una serie estacionaria, se pasa a establecer la forma del modelo a emplear. Esto implica calcular la autocorrelación y autocorrelación parcial de varios modelos ARIMA, y seleccionar un modelo tentativo que mejor describa la serie.(Mondal et al., 2014)

#### Paso 2: Estimación de modelos

Luego de identificar el modelo tentativo, se procede a estimar los parámetros del modelo ARIMA. Esta estimación se lleva a cabo minimizando la suma de los cuadrados de los desajustes de ajuste mediante un procedimiento no lineal llamado mínimos cuadrados.

#### Paso 3: Evaluación del modelo

En este paso se busca determinar si el modelo estimado es bueno si se acomoda a los datos y si sus residuales no pueden usarse para mejorar los pronósticos y deben de ser aleatorios. (Santiago de la Fuente, 2012)

Es necesario llevar a cabo un análisis de regresión sobre los residuos de un modelo ARIMA. Los histogramas y los gráficos de probabilidad normal son herramientas útiles para validar la normalidad de los datos, mientras que la gráfica de la secuencia de tiempo es útil para detectar datos atípicos.

Las autocorrelaciones residuales deben ser limitadas y pertenecer al intervalo de  $\pm 2 / \sqrt{n}$  alrededor de cero. La presencia de autocorrelaciones residuales significativas en retardos cortos o estacionales indica que el modelo no es apropiado y necesita ajustes. Además, estas autocorrelaciones residuales deben concordar con aquellas generadas por errores aleatorios.

Además, las autocorrelaciones residuales deben ser consistentes con las generadas por errores aleatorios.

La prueba de chi-cuadrado, fundamentada en la estadística de Ljung-Box Q, se empleará para evaluar la idoneidad del modelo. Este test evalúa las autocorrelaciones como un conjunto. La estadística de prueba Q es utilizada.

$$Q_m = n(n + 2) \sum_{k=1}^m \frac{r_k^2(e)}{n - k}$$

Donde:

k=el retraso del tiempo.

$r_k(e)$ = La autocorrelación residual del retardo k.



$m$ = Número de retrasos de tiempo que tienen que ser estimados.

$n$ = Número de residuales.

$r$ = Parámetros estimados.

Si el valor  $p$  que está asociado con la estadística  $Q$  es ( $p < 0.05$ ) se considera que el modelo no es el adecuado en este caso se debe de examinar un modelo nuevo o modificarlo y proseguir con el análisis hasta que se encuentre el modelo adecuado.

#### Paso 4: Pronósticos del modelo

Una vez que se tiene el modelo adecuado se llevan a cabo los pronósticos para uno o varios periodos en el futuro, de igual manera se pueden construir intervalos de predicción.

Mientras más largo sea el tiempo guía del pronóstico mayor será la predicción.

#### 2.2.16. Test de Shapiro – Wilk

Según (Carmona Arce & Carrion Rosales, 2015), el test de Shapiro-Wilk se utiliza para evaluar si una muestra sigue una distribución normal, siendo especialmente útil para muestras pequeñas. Así mismo, menciona que los pasos básicos para llevar a cabo este test son:

Establecer las hipótesis:

- Hipótesis nula ( $H_0$ ): La muestra sigue una distribución normal.
- Hipótesis alternativa ( $H_1$ ): La muestra no sigue una distribución normal.

Calcular el estadístico W de Shapiro-Wilk: Este estadístico se obtiene mediante una fórmula que utiliza los coeficientes de la distribución normal y los datos de la muestra.

Obtener el valor p:

- El valor p asociado con el estadístico W ayuda a tomar una decisión sobre las hipótesis.

Tomar una decisión:

- Si el valor p es menor que el nivel de significancia (normalmente 0.05), se rechaza la hipótesis nula, concluyendo que la muestra no sigue una distribución normal.
- Si el valor p es mayor que el nivel de significancia, no se rechaza la hipótesis nula, sugiriendo que la muestra podría seguir una distribución normal.

### 2.2.17. Test de Dickey – Fuller Aumentado (ADF)

Este test es uno de los más usados para determinar la estacionariedad en una serie temporal. Por lo cual este es un contraste de “No estacionariedad”. Aunque el decaimiento lineal aproximado de la ACF de la muestra suele tomarse la presencia de autocorrelaciones significativas en los residuos puede indicar que la serie temporal subyacente no es estacionaria y necesita ser diferenciada. También puede servir para evaluar la evidencia de no estacionariedad en el proceso de generación de datos (Cryer & Chan, 2008) Esto puede hacerse mediante pruebas de hipótesis. Consideremos el modelo:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t \quad t = 1, 2, \dots$$

Los tests del Dickey-Fuller Aumentado (ADF) se originaron bajo la premisa restrictiva de que la serie se ajusta a un proceso AR(1). Sin embargo, la serie podría seguir procesos más complejos como ARMA(p,q). Cualquier proceso ARMA(p,q) puede aproximarse a un AR(p) con el grado de precisión requerido. (González Casimiro, 2009), suponiendo el modelo:

$$\Delta Y_t = \alpha_0 + \gamma Y_{t-1} + \sum_{i=2}^B B_i \Delta Y_{t-i+1} + \varepsilon_t$$

$H_0: \gamma = 0$  raíz ununitaria (proceso no estacionario).

$H_1: \gamma < 0$  no hay raíz ununitaria (proceso estacionario).

El “test de Dickey Fuller” es una evaluación que busca determinar la presencia de raíces unitarias. En el caso del “Test de Dickey Fuller Aumentado”, se toma en cuenta la correlación del error. Para aplicar esta prueba, se parte del modelo aleatorio puro y se resta el término autorregresivo de ambos lados de la ecuación.

$$y_t - y_{t-1} = \phi y_{t-1} - y_{t-1} + u_t$$

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

Dónde:

$$\Delta y_t = y_t - y_{t-1}$$

$$\gamma(\phi - 1)$$

Utilizando una raíz unitaria  $\phi = 1$  y un parámetro de  $\gamma = 0$  se puede utilizar el Método de (MCO) en las primeras etapas del modelo, sin incluir una constante. El coeficiente estimado del término autorregresivo puede ser sometido



a la prueba t convencional para evaluar su significancia estadística y contrastar la hipótesis nula de presencia de raíz unitaria frente a la hipótesis alternativa de estacionariedad:

$$H_0: \gamma = 0 \text{ por consiguiente } \phi = 1$$

$$H_A: \gamma = 0 \text{ por consiguiente } \phi < 1$$

El Test de Dickey Fuller Aumentado se utiliza para evaluar la estacionariedad de una serie, en razón a que su hipótesis nula implica la existencia de una raíz unitaria en el proceso de la serie bajo análisis.

#### 2.2.18. Estadístico de Box – Pierce

Según Novales (2016); Box y Pierce (1970) sugirieron un conjunto de contrastes basados en un estadístico para capturar de manera adecuada el concepto de un proceso de camino aleatorio, donde todos los coeficientes de autocorrelación son nulos:

$$Q_m \equiv T \sum_{k=1}^m p_k^2$$

Bajo la hipótesis nula  $H_0: p_1 = p_2 = \dots = p_m = 0$ , (RW1), (7) implica que:

$$Q_m \equiv T \sum_{k=1}^m p_k^2 \text{ se distribuya como una } \chi_m^2.$$

Ljung y Box (1978) Se planteó una corrección para el caso de muestras infinitas:

$$Q'_m \equiv T(T + 2) \sum_{k=1}^m \frac{p_k^2}{T - k}$$

Al elevar al cuadrado los coeficientes de autocorrelación, el contraste obtiene capacidad para detectar varias alternativas. No obstante, la determinación del orden  $m$  puede ser ambigua y podría influir en los resultados del contraste. Si se elige un  $m$  pequeño, es posible que no se detecte la autocorrelación de orden superior. Por otro lado, si se selecciona un  $m$  grande, el contraste podría perder capacidad de detección debido a la acumulación de autocorrelaciones que no son significativas. En caso de contar con una alternativa, sería beneficioso diseñar un contraste que posea mejores propiedades estadísticas. (González Casimiro, 2009)

#### **2.2.19. Criterio de Información de Akaike (AIC):**

Caballero Díaz (2011) menciona que el criterio denominado como AIC es una medida utilizada para la selección de modelos que equilibra la bondad de ajuste del modelo con la simplicidad del mismo. Su objetivo es encontrar el modelo que mejor se ajuste a los datos observados, considerando tanto la bondad del ajuste como la complejidad del modelo, Fue propuesto por Hirotugu Akaike en 1974 y se calcula mediante la fórmula:

$$AIC = -2\log(L) + 2k$$

Donde:

- $L$  es la función de verosimilitud del modelo, que indica cuán bien se ajusta el modelo a los datos observados.
- $k$  es el número de parámetros estimados en el modelo.



El AIC penaliza los modelos más complejos añadiendo  $2k$  al valor  $-2\log(L)$ . Por lo tanto, un valor menor de AIC indica un mejor ajuste del modelo a los datos observados, considerando la complejidad del modelo.

El AIC ofrece un método sistemático para evaluar y comparar modelos estadísticos. Al calcular el AIC para varios modelos, aquel con el valor más bajo se interpreta como el más adecuado, ya que logra un equilibrio ideal entre ajuste y complejidad, facilitando así la selección del modelo más apropiado para los datos analizados.

### **2.2.20. Criterio de información Bayesiano (BIC)**

El Criterio de Información Bayesiano (BIC), también conocido como Criterio de Información de Schwarz, es una medida utilizada en estadística y modelado para la selección de modelos. Fue propuesto por Gideon Schwarz como una alternativa al Criterio de Información de Akaike (AIC). A diferencia del AIC, el BIC penaliza más fuertemente los modelos complejos, especialmente en muestras pequeñas. (Fernández Casal et al., 2021)

La fórmula para el BIC es:

$$\text{BIC} = -2\log(L) + p \log(n)$$

Donde:

L es la función de verosimilitud del modelo.

p es el número de parámetros en el modelo.

n es el tamaño de la muestra.



El BIC incorpora un término adicional  $p \log(n)$  que penaliza modelos con más parámetros de manera más estricta que el AIC. Esta penalización aumenta con el tamaño de la muestra  $n$ , lo que hace que el BIC sea particularmente útil en contextos donde se desea evitar el sobreajuste del modelo a datos limitados.

A diferencia del Criterio de Información de Akaike (AIC), el Criterio de Información Bayesiano (BIC) penaliza de forma más rigurosa los modelos complejos, especialmente en muestras pequeñas. Este enfoque ayuda a prevenir el sobreajuste al favorecer modelos más simples que tienen una mejor capacidad de generalización hacia datos no observados.

## **2.3. MARCO CONCEPTUAL**

### **2.3.1. Mortalidad**

La mortalidad estudia la frecuencia de muertes producidas dentro de una población, área geográfica y período determinado.

La mortalidad, en un sentido amplio, se refiere a la cualidad mortal inherente a todos los seres vivos. De manera más específica, se trata del número de fallecimientos ocurridos durante un lapso específico en una región geográfica. En el ámbito de la demografía, el estudio de la mortalidad se integra con los conceptos de natalidad y migración, cuyas tasas combinadas ofrecen una visión completa del crecimiento poblacional. (Ministerio de Salud del Perú, 2009)

### **2.3.2. Mortalidad por sexo**

Según Raguz (2015) El sexo se refiere a cómo la sociedad y la cultura interpretan las características biológicas, características anatómicas, fisiológicas y cromosómicas que definen la identidad de las personas como mujer u hombre. Por



lo tanto, se basa en cómo se entienden estas diferencias biológicas en un contexto social y cultural.

La mortalidad por sexo se define como la muerte que se da dentro de una población diferenciado por el sexo ya sea femenino o masculino debido a diferentes causas, quiere decir que es el índice de muertes ya sea masculinas o femeninas.(Sanchez Perez, 2021)

### **2.3.3. Violencia de género y violencia contra la mujer**

#### **2.3.3.1. Violencia de género**

Este tipo de violencia afecta a individuos, independientemente de su sexo, debido a las expectativas y roles asociados a su identidad de género en la sociedad. Esta forma de violencia surge de diversas influencias en la personalidad de cada individuo en su entorno social. El término "género" abarca el conjunto de normas y roles que definen lo que significa ser hombre o mujer en una comunidad específica. Va más allá del aspecto biológico y establece un sistema de jerarquías que impacta a todas las personas en la sociedad. (Bodelón, 2014)

#### **2.3.3.2. Violencia contra la mujer**

Según la Convención de Belém do Pará (1994) la violencia contra las mujeres engloba todo tipo de acción que les cause daño físico, sexual, psicológico o incluso la muerte, simplemente por ser mujeres. Este tipo de violencia puede ocurrir tanto en entornos públicos como privados, y su propósito es dominar, controlar o infundir miedo a las mujeres debido a su condición de género.



- Aquella que acontezca en el entorno familiar o convivencia doméstica, así como en cualquier tipo de relación interpersonal, sin importar de si el agresor convive o ha convivido previamente con la mujer. Esto abarca una amplia gama de acciones, incluyendo, la violación, maltrato físico o psicológico, y abuso sexual entre otros.
- La violencia comunitaria, llevada a cabo por cualquier persona, abarca una variedad de acciones como el abuso sexual, la explotación sexual, los secuestros dentro de las instituciones públicas o privadas.

#### Definición de violencia contra la mujer según la ONU

“Cualquier acto de violencia hacia las mujeres por su condición de género que tenga o pueda tener como resultado un daño o sufrimiento físico, sexual o psicológico para la mujer, incluyendo las amenazas, la coacción o la privación de la libertad, tanto si se produce en la vida pública o la privada”. (ONU, 1993)

Definición de violencia contra la mujer según la Organización Mundial de la Salud (OMS),

“La aplicación intencional de la fuerza o el poder físico, ya sea de manera real, o como amenaza, contra uno mismo, otra persona, un grupo o una comunidad, que provoque o tenga muchas posibilidades de causar, lesiones, muerte, daños psicológicos, trastornos del desarrollo o privaciones”.(OMS, 2024)



#### **2.3.4. Política penal del delito de feminicidio**

En la legislación peruana, “el feminicidio siempre es un delito doloso o intencional, donde no es admisible la acción involuntaria o culpa”. (Ley que modifica el Código Penal y el Código de los Niños y Adolescentes - LEY - N° 30819 - CONGRESO DE LA REPUBLICA, 2018)

##### **“Artículo 108-B.- Feminicidio”**

Cualquier individuo que cause la muerte de una mujer debido a su género enfrentará una pena de prisión no menor de quince años en varios casos que estipula el Artículo 108-B.

A esto se suma el “Artículo 121-B.- Lesiones graves por violencia contra las mujeres e integrantes del grupo familiar”

Por otra parte, el “Artículo 122-B.- Agresiones en contra de las mujeres o integrantes del grupo familiar” sostiene:

Quien genere lesiones corporales a una mujer debido a su género o a miembros del grupo familiar, estas que llegarían a afectar tanto al ámbito psicológico, cognitivo o conductual, así mismo en su artículo 108-B, sostiene que esta puede ser condenado a una pena de prisión que oscilará entre uno y tres años, además de la inhabilitación según lo establecido en el artículo 36.

En el caso de que se presenten ciertas circunstancias agravantes en los escenarios mencionados en el primer párrafo, la pena será de no menos de dos años ni más de tres años:



- Si se utiliza cualquier tipo de arma, objeto que ponga en peligro a la víctima.
- Si el acto se desarrolla con ensañamiento o alevosía.
- Si el acto se comete cuando la víctima está embarazada.
- Si la víctima es menor de edad o presenta alguna discapacidad o es adulta mayor y el agresor se sirve de esta situación.

### **2.3.5. Femicidio**

Existe una gran cantidad de definiciones sobre el Femicidio, como por ejemplo (Russell & Harmes, 2006) en su libro “Femicidio: Una Perspectiva Global” El concepto de femicidio abarca el crimen de odio dirigido específicamente contra las mujeres, englobando diversas formas de violencia que pueden desembocar en asesinatos o incluso en suicidios femeninos. Esta distinción se hace para diferenciar los términos y se emplea el término "femicidio" Para hacer referencia al grupo de crímenes de lesa humanidad que abarcan no solo los homicidios, sino también las desapariciones, los secuestros sean de niñas y mujeres.

De acuerdo a lo señalado por la (RAE, s. f.) el femicidio se describe como el crimen que consiste en causar la muerte de una mujer en condiciones específicas, como relaciones sentimentales o conyugales, la existencia de confianza con la víctima, conexiones laborales, subordinación o superioridad, con el fin de satisfacción sexual, encubrimiento de violación, o por cualquier motivo derivado de su condición de mujer, todo esto dentro de un contexto de relaciones marcadas por desigualdades de poder.





### **2.3.6. Características del Femicidio**

Es importante conocer cuáles son las características principales del feminicidio, lo que nos ayudará de forma efectiva a identificar un caso de violencia, como lo menciona Moreno y Lizárraga (2017) que en su mayoría los feminicidios reportados implican a una pareja anterior o actual de la víctima, y en gran parte de los casos, el ciclo de feminicidio se inicia con formas de violencia psicológica, física y sexual. Ejemplos de esto incluyen el uso de lenguaje agresivo en el hogar, abuso emocional, intimidación, golpes, coerción sexual y situaciones en las que las mujeres se ven privadas de poder o recursos en comparación con sus parejas. Estos comportamientos son considerados socialmente como una grave falta de respeto, una minimización y una desvalorización de la mujer, reflejando el enorme poder y control que se ejerce sobre la víctima dentro de la relación.

La edad de las mujeres víctimas de feminicidio es un factor importante, ya que diversas investigaciones han revelado que el 55% de las víctimas se encontraban en el rango de edad de 16 a 35 años en el momento de su muerte. Esta estadística sugiere que las mujeres en esta franja etaria son especialmente vulnerables a ser víctimas de feminicidio. Además, indica que, a pesar de que estas mujeres podrían tener la capacidad de denunciar la violencia que están experimentando, muchas de ellas no lo hacen debido al miedo y las amenazas que reciben por parte de sus agresores, lo que refleja el profundo impacto psicológico y emocional que sufren. (López & Rissi, 2008)

### **2.3.7. Causas del Femicidio**

Es importante aclarar que no hay causas por las cuales se debería cometer Femicidio y que la vida de todas las personas es importantes e iguales, pero hay



situaciones que se contempla en el feminicidio según la ley N° 30364 realizado por el Gobierno del Perú (2015). Las cuales son:

### **2.3.7.1. Causado por Violencia Familiar**

Este tipo de feminicidio se limita a ocurrir dentro del ámbito familiar, que incluye tanto la familia nuclear como la familia conyugal. Para abordar y resolver esta forma de feminicidio, se recurre a la aplicación de la Ley N° 30364, que tiene como objetivo “prevenir, sancionar y erradicar la violencia contra las mujeres y los miembros del grupo familiar”. Dicha ley aborda la violencia familiar desde dos perspectivas: la violencia dirigida específicamente contra las mujeres y la violencia dirigida hacia los integrantes del grupo familiar en su conjunto.

La Ley N° 30364, en su artículo 5, establece la definición de violencia contra las mujeres como cualquier acto o comportamiento que resulte en la muerte, lesiones o sufrimiento físico, sexual o psicológico, simplemente por el hecho de ser mujer. Esta normativa precisa que se considera violencia contra las mujeres si:

Las que tengan lugar en el ámbito familiar o en relaciones interpersonales donde el agresor haya compartido o comparta el mismo lugar de residencia que la mujer, cubriendo acciones como violación, maltrato, abuso sexual, entre otros.

Aquellas que sucedan en la comunidad, llevadas a cabo por cualquier persona, e involucren acciones como violación, abuso sexual, tortura, trata de personas, explotación sexual, secuestro y acoso sexual en el entorno laboral, así como en escuelas, centros médicos u otros lugares.



El artículo 6 de la Ley mencionada también delinea “la violencia contra los integrantes del grupo familiar como cualquier acción o conducta que resulte en la muerte, lesiones o sufrimiento, y que tenga lugar dentro de una relación de responsabilidad entre los integrantes del grupo familiar. Esta disposición se enfoca especialmente en la protección de niñas, niños, adolescentes, adultos mayores y personas con discapacidad dentro del entorno familiar”.(Gobierno del Perú, 2015)

#### **2.3.7.2. Causado por Constreñimiento**

Este conjunto de acciones ilícitas se engloba en los casos de coacción, hostigamiento o acoso sexual. Por otra parte, la coacción en el marco del feminicidio está establecido en el artículo 151 del Código Penal, que conceptualiza la coacción como la acción de obligar a alguien a hacer algo que la ley no prescribe mediante amenazas o violencia. Es importante destacar que la coacción no constituye el medio para cometer el feminicidio, sino que actúa como un estado previo al homicidio contra la mujer motivado por razones de género.

En el contexto del hostigamiento que también se puede interpretarse como proveniente de un entorno tanto interpersonal como laboral.

Dentro del contexto del acoso sexual, se refiere al homicidio en perjuicio de la mujer debido a su género, cuando el agresor la estaba acosando, ya sea en el lugar de trabajo u otro entorno, ejerciendo presión sobre ella con el objetivo de obtener gratificación sexual. En este escenario, si el agresor no logra su objetivo, es decir, que la mujer no cede



a su acoso, puede llegar a asesinarla por no aceptar que una mujer lo rechace.

Este tipo de comportamiento es una manifestación extrema del patriarcado y la misoginia, donde el agresor busca controlar y someter a la mujer, y ante el rechazo, recurre a la violencia extrema para mantener su supuesto "dominio".

### **2.3.7.3. Causado por dominio o influjo sobre la víctima**

En esta clasificación, el perpetrador comete el crimen al aprovechar su poder, confianza, posición dominante o relación de subordinación con respecto a la víctima. A diferencia del feminicidio en contextos de violencia familiar, aquí la autoridad no proviene del ámbito familiar, sino de un entorno externo, como el laboral o grupal. Por otra parte, el abuso de poder generada por la confianza u otra posición de autoridad implica el ejercicio arbitrario de ese poder, que puede manifestarse en relaciones laborales, ya sea en el sector privado o público.

Por ejemplo, un ejecutivo de una empresa que puede ser el superior jerárquico de la víctima y que, sin mediar acoso sexual, causa la muerte debido a su género. Otro ejemplo sería el líder de un grupo, pandilla o comunidad que, aprovechándose de su posición de liderazgo por motivos de género, provoca intencionalmente la muerte de una mujer miembro del grupo.(Gobierno del Perú, 2015)



#### **2.3.7.4. Causado por el estado de gestación de la víctima**

Este tipo de violencia hace referencia al acto de homicidio cometido contra una mujer debido a su género, especialmente cuando está embarazada, lo que implica que la mujer está llevando en su vientre a otro ser humano.

El agresor mata a la fémina por no querer que el bebé nazca, esto representa una alarma social grande porque este hecho está segando dos vidas humanas, Sin embargo, para que se haga efectivo el delito, el sujeto tenga conocimiento que la mujer se encontraba gestando. Caso contrario si el agresor mata a la mujer sin tener conocimiento de su gestación y estando la víctima en las primeras etapas del embarazo, su conducta no sería condenada como agravante, sino como en el tipo feminicidio de tipo base.

#### **2.3.8. Consecuencias del Feminicidio**

El feminicidio acarrea consecuencias devastadoras que incluyen la pérdida de vidas femeninas de diversas edades, dejando a familias traumatizadas, menores sin padres y desprotegidos, y creando un ambiente en el que los niños aprenden patrones de violencia que pueden repetir en el futuro, alimentando así un ciclo pernicioso. Esta tragedia no solo se limita al impacto inmediato de la pérdida de vidas, sino que también genera repercusiones de largo alcance en las futuras generaciones.(Bernáldez Jaimes et al., 2023)



### **2.3.9. Factores determinantes del feminicidio**

El feminicidio se produce cuando una mujer pierde la vida como resultado de actos de coacción, hostigamiento o acoso sexual. Los actos de coacción implican obligar a alguien a hacer algo en contra de su voluntad. Por otro lado, el hostigamiento se refiere a molestar de forma persistente a alguien con el objetivo de, por ejemplo, provocar su renuncia en el trabajo o inducirlo a realizar acciones a favor del acosador. En cuanto al acoso sexual, se refiere a solicitar de manera insistente que una mujer acceda a tener relaciones sexuales. Estas formas de violencia reflejan un grave desprecio por la autonomía y la dignidad de la mujer, y lamentablemente pueden tener consecuencias trágicas. Es fundamental abordar estas conductas para prevenir la pérdida de vidas femeninas y garantizar la seguridad y el respeto hacia los demás.

Según Lopez y Jhon (2021) existen tres factores que son los más comunes para que ocurra el feminicidio los cuales son:

#### **Celos**

Los celos son un factor que puede llevar tanto a hombres como a mujeres a perder el control emocional, ya que esta emoción tiene el poder de perturbar y desviar nuestra atención de la realidad, llevándonos a actuar de manera impulsiva y emocional en muchas ocasiones. Es una experiencia emocional intensa que puede nublar nuestro juicio y hacernos perder la perspectiva racional de las situaciones. Los celos pueden generar una serie de reacciones y comportamientos irracionales que pueden tener repercusiones negativas en nuestras relaciones interpersonales y en nuestro bienestar emocional. Es importante aprender a gestionar y controlar esta emoción para mantener relaciones saludables y una vida



emocional equilibrada. Por ello, un motivo principal detrás de la perpetración de feminicidios es la presencia de celos, usualmente expresados por esposos, novios o exparejas. Estos individuos tienden a ver a las mujeres como posesiones y actúan movidos por el deseo de ejercer control sobre ellas. Quieren demostrar su dominio ante la sociedad.

### Cultura Machista

El proceso de socialización comienza desde la niñez, sobre todo en el ambiente del hogar, donde se establecen las primeras percepciones de género que diferencian a los niños de las niñas. Los roles de género y las expectativas sociales se transmiten y internalizan desde la infancia, influenciando la manera en que los niños y niñas se perciben a sí mismos y a los demás. En la familia, se puede observar un conjunto de creencias y actitudes arraigadas que perpetúan normas de género tradicionales y estereotipadas, moldeando así las identidades de género de los niños desde una edad temprana “Disfruta ahora la vida, que estás soltera, porque el día que te cases, ese día dejas de ser libre, hasta ese día dura la felicidad”, también “Tienes que cocinar y atender bien al marido para que te trate bien” Estos son pensamientos machistas provienen luego de unión conyugal.

### Violencia innata del ser humano

La violencia, lamentablemente, parece ser una característica inherente a la naturaleza humana, aunque hasta ahora no se haya realizado una investigación exhaustiva sobre este aspecto para abordar y reducir esta situación de manera efectiva. Esta tendencia se observa de manera recurrente, especialmente en el comportamiento masculino, donde los hombres tienden a manifestar actitudes violentas más frecuentemente que las mujeres. Esta incapacidad de controlar los



impulsos violentos es una realidad compleja y difícil de abordar. Parece que la falta de educación juega un papel importante en este problema, ya que una educación deficiente dificulta la capacidad de diferenciar los niveles de agresión hacia la pareja. Es fundamental profundizar en la comprensión de este fenómeno y en la aplicación de estrategias efectivas de prevención y educación para promover relaciones saludables y libres de violencia.





## CAPÍTULO III

### MATERIALES Y MÉTODOS

#### 3.1. UBICACIÓN GEOGRÁFICA DEL ESTUDIO

Este estudio de investigación se lleva a cabo en el Perú, un país con una población de 31.4 millones de habitantes y cuya ciudad capital es Lima, que se extiende sobre 1 285 215,6 km<sup>2</sup>, siendo reconocido como uno de los 19 países más extensos del mundo. Geográficamente, el Perú se compone de tres regiones principales: la Costa, la Sierra y la Selva.

#### 3.2. PERIODO DE DURACIÓN DE ESTUDIO

El trabajo de investigación tuvo un proceso de recopilación de información de los años 2010 al 2021 del total de feminicidios por año en el Perú, con el propósito de evaluar las futuras proyecciones del Feminicidio.

#### 3.3. PROCEDENCIA DEL MATERIAL UTILIZADO

Al tener ya identificado el diseño de la investigación y la muestra acorde al problema de estudio, para el proceso siguiente se recolecta todos los datos extraídos del Portal Estadístico Del Programa Nacional AURORA que son necesarios para llevar a cabo el pronóstico.

#### 3.4. POBLACIÓN Y MUESTRA DE ESTUDIO

##### 3.4.1. Población

La población para el siguiente estudio está constituida por el número total de defunciones a causa del Feminicidio, que se registra desde la creación en el año



2001 del “Programa Nacional para la Prevención y Erradicación de la Violencia contra las Mujeres y los Integrantes del Grupo Familiar” – AURORA.

### **3.4.2. Muestra**

El criterio de selección de la muestra se basó en un muestreo no probabilístico, a criterio del investigador, debido a que los datos disponibles desde la creación del programa hasta el año 2009 no son completos ni concretos, lo cual afectaría negativamente la identificación del modelo. Por tal motivo, la muestra está representada por los 144 meses comprendidos desde enero de 2010 hasta diciembre de 2021, abarcando el número total de feminicidios en el Perú registrados en los informes estadísticos del programa nacional AURORA.

## **3.5. DISEÑO METODOLÓGICO**

### **3.5.1. Diseño de la Investigación**

El diseño de la presente es no-experimental porque no se da un control experimental en las variables de estudio, ya que no se realizó el proceso de aleatorización y nuestro interés es analizar cambios a través del tiempo de la variable. Entonces, se utilizan diseños longitudinales, ya que los datos se recopilan a lo largo del tiempo en intervalos específicos, lo que permite inferir sobre el cambio y sus consecuencias.

### **3.5.2. Enfoque de Investigación**

Este estudio de investigación adopta un enfoque cuantitativo, ya que implica la recopilación y análisis de datos numéricos, específicamente el recuento de feminicidios. Este método se considera idóneo para identificar tendencias y realizar predicciones basadas en los datos recopilados.



### **3.5.3. Tipo de Investigación**

Este tipo de investigación se centra en brindar una descripción detallada y una exploración de los fenómenos en diferentes situaciones y contextos. Por lo tanto, el objetivo principal de este estudio es describir la evolución mensual de los casos de feminicidio en los últimos 10 años, lo que permitirá comprender mejor la magnitud y las tendencias de este grave problema social.

### **3.5.4. Método de Investigación**

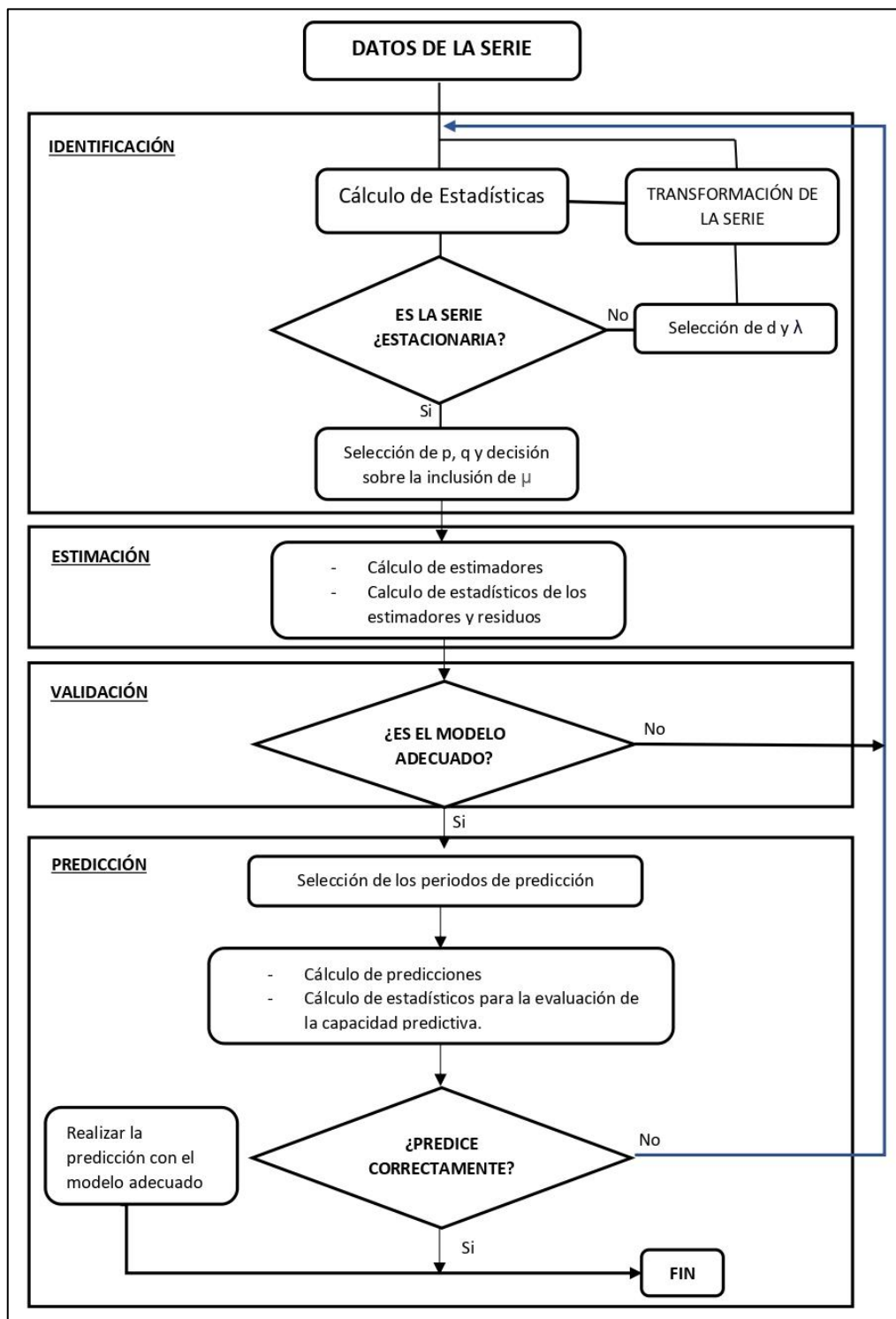
La metodología seleccionada fue la de “Box-Jenkins”, conocida como la metodología ARIMA, que se basa en el análisis de las propiedades probabilísticas o estocásticas de las series temporales.

#### **METODOLOGÍA DE BOX-JENKINS**

Para analizar el número de feminicidios en el Perú, se ha decidido utilizar la “metodología de Box-Jenkins” (también conocida como la teoría de Wiener-Kolmogorov), la cual implica un proceso dividido en cuatro etapas.

**Figura 10**

*Diagrama de flujo de la metodología Box – Jenkins.*





### **3.6. PROCEDIMIENTO DE RECOLECCIÓN DE DATOS**

Los datos fueron extraídos del Portal Estadístico del Programa Nacional AURORA.

Las cifras de los feminicidios, se tabuló en forma mensual de los años del 2010 – 2021.

#### **3.6.1. Técnicas de recolección de datos**

La técnica utilizada en el presente trabajo de investigación es de tipo documental, el cual fue extraído de la base del Portal Estadístico del Programa Nacional AURORA, la base de datos fue alimentada a través de los Centros de Emergencia Mujer (CEM) que son servicios públicos especializados y gratuitos, enfocado en las víctimas de violencia contra la mujer e integrantes del grupo familiar, como también El Comité Estadístico Interinstitucional de la Criminalidad (CEIC), fue creado mediante “Decreto Supremo N° 013-2013-JUS, presidida por el Instituto Nacional de Estadística e Informática (INEI), e integrada por representantes del Poder Judicial (PJ), Ministerio Público (MP), Ministerio del Interior (MININTER), Policía Nacional de Perú (PNP), Ministerio de Justicia y Derechos Humanos (MINJUSDH)”.

Además, para establecer la cifra oficial de feminicidios, el CEIC incluye a todas las víctimas cuya investigación está al menos en la etapa de investigación fiscal preliminar o en una fase más avanzada del proceso de judicialización.



### 3.7. VARIABLES

**Tabla 2**

*Operacionalización de variables*

VARIABLE	DEFINICIÓN	DIMENSIÓN	INDICADOR	ESCALA
FEMINICIDIO	Homicidio hacia las mujeres por situaciones que se encuentran relacionadas al odio de género, celos, machismo, conductas psicópatas, entre otros ocasionando la muerte de una mujer ya sea de manera privada o pública y que exista o no algún tipo de relación entre el agresor y la víctima.	Niveles estadísticos de Femicidio Comportamiento del femicidio año tras año	Estadísticas de homicidios contra las mujeres Incremento de feminidios en el Perú	Razón



## CAPÍTULO IV

### RESULTADOS Y DISCUSIÓN

En esta última etapa se muestra los resultados que fueron obtenidos de la información de la serie histórica de feminicidios mensuales en el Perú, durante el periodo 2010 al 2021.

Los resultados de esta investigación se obtuvieron mediante el software RStudio. Además, para el desarrollo de esta investigación aplicaremos la metodología Box Jenkins, la cual consiste en desarrollar las cuatro fases de la metodología (identificación, estimación, validación y pronóstico), con la finalidad de encontrar el mejor modelo ARIMA para pronosticar los casos de feminicidio en el Perú para el año 2022.

En la **Tabla 3** se muestra la información del número de defunciones a causa del Feminicidio, que se registra en el informe estadístico de Feminicidios en el Perú, del “Programa Nacional para la Prevención y Erradicación de la Violencia contra las Mujeres y los Integrantes del Grupo Familiar” – AURORA, para este análisis se cuenta con información recogida de 12 periodos (años) desde el año 2010 al 2021, haciendo un total de 144 observaciones.

**Tabla 3***Datos De Información*

MESES	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021
<b>Enero</b>	13	13	7	11	7	8	6	8	10	15	20	17
<b>Febrero</b>	10	7	6	6	10	9	8	12	12	14	13	9
<b>Marzo</b>	7	8	8	7	11	5	9	9	11	13	5	10
<b>Abril</b>	14	6	3	8	11	8	8	5	10	13	10	10
<b>Mayo</b>	7	3	7	12	8	10	10	10	19	11	11	12
<b>Junio</b>	5	1	7	11	9	5	12	14	8	17	10	18
<b>Julio</b>	13	5	8	8	10	9	17	13	12	13	9	12
<b>Agosto</b>	11	7	10	15	1	10	14	11	11	18	13	14
<b>septiembre</b>	6	8	7	7	2	7	15	12	10	7	11	6
<b>Octubre</b>	14	6	1	16	8	3	8	5	16	13	12	8
<b>noviembre</b>	12	16	11	19	9	12	9	10	13	15	12	9
<b>diciembre</b>	9	13	8	11	10	9	8	12	17	17	5	11
<b>TOTAL</b>	121	93	83	131	96	95	124	121	149	166	131	136

Nota: Informe estadístico de feminicidios en el Perú.

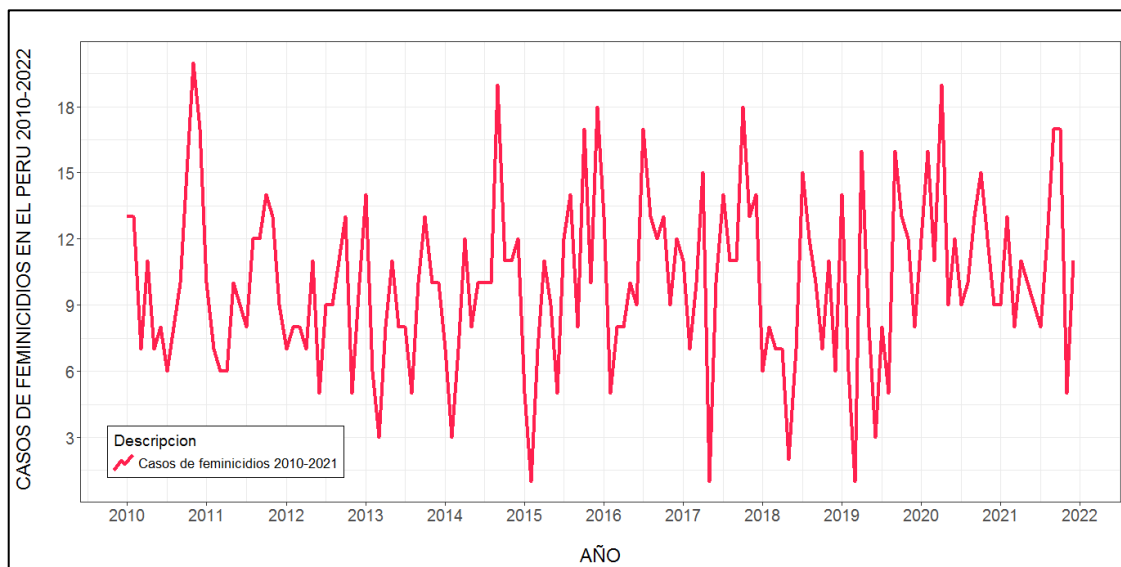
#### 4.1. IDENTIFICACIÓN DEL MODELO ARIMA

Se tiene en la **Figura 11** la serie casos de feminicidio en el Perú periodo 2010-2021 graficados en un periodograma, realizando un análisis visual a la figura, podemos afirmar que, la serie no presenta tendencia, al contrario, tiene un comportamiento aleatorio propio de las series estacionarias, es decir que, los datos fluctúan alrededor de su media y por lo tanto da algunos indicios de que podríamos estar frente a una serie estacionara.



## Figura 11

Gráfico de la serie casos de feminicidio 2010-2021



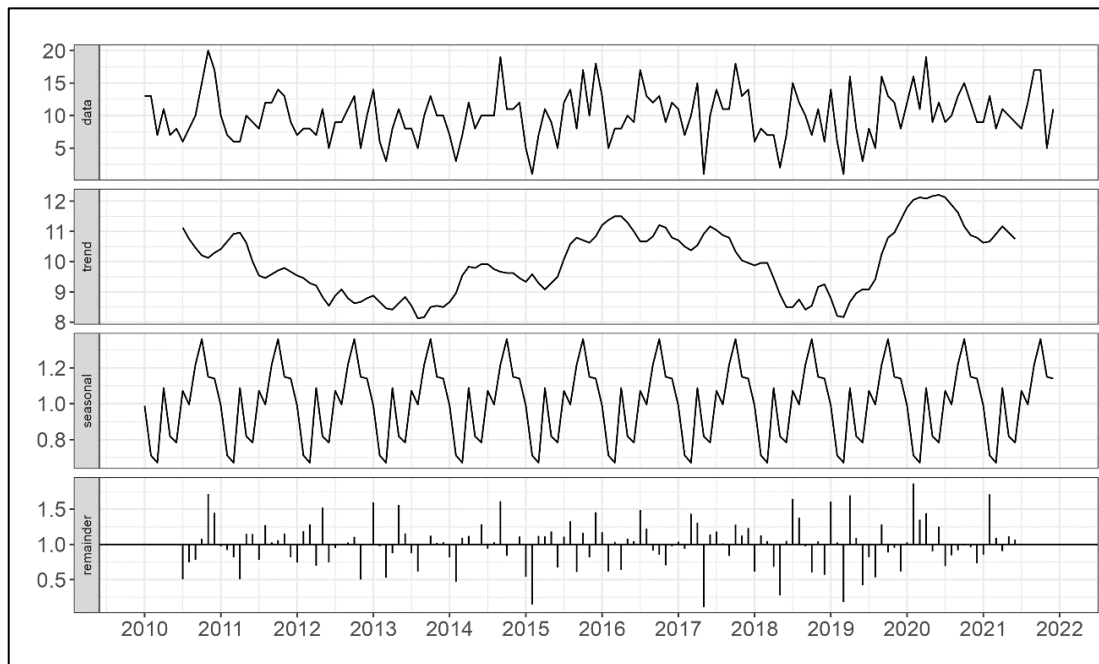
Nota: elaboración propia del autor.

La descomposición de una serie temporal es uno de los pasos que propone Box-Jenkins en su metodología, para analizar a través de sus cuatro componentes (tendencia, estacionalidad, ciclo, error) la estacionariedad o no de una serie temporal.

Se muestra en la **Figura 12** la descomposición multiplicativa de la serie casos de feminicidios 2010-2021, este procedimiento muestra a la serie original dividido en sus cuatro compontes para su análisis; la serie no presenta tendencia tal como se observa en la **Figura 12** (trend), con respecto a la componente estacional la serie contiene estacionalidad en distintos periodos con una duración larga (seasonal), además la serie presenta patrones que se vuelven a repetir en periodos largos a esto le denominamos ciclo, y por último la serie contiene aleatoriedad en sus datos con fluctuaciones que van de arriba hacia abajo con respecto a su media.

**Figura 12**

*Descomposición multiplicativa de la serie casos de feminicidios 2010+-2021*



Nota: Descomposición multiplicativa a la serie original

Fuente: Elaboración propia de los autores

En la **Tabla 4** se muestra las estadísticas descriptivas de la serie en estudio, la información revela que, el número máximo de feminicidios ocurridos en el Perú sucedió en el mes de enero-2020 con un número máximo de 20 casos de feminicidio, seguido de un mínimo de 1 caso ocurridos entre los meses de junio-2011, octubre-2012 y agosto-2014; los datos varían con respecto a su promedio en 37.97% que viene hacer la desviación estándar.

**Tabla 4**

*Estadísticas descriptivas de la serie casos de feminicidio en el Perú 2010-2021*

Datos	Maximo	Mínimo	Media	Moda	Mediana	Desviación Estandar	Coefficiente De Variación
144	20	1	10.04167	8	10	3.813621	37.97797

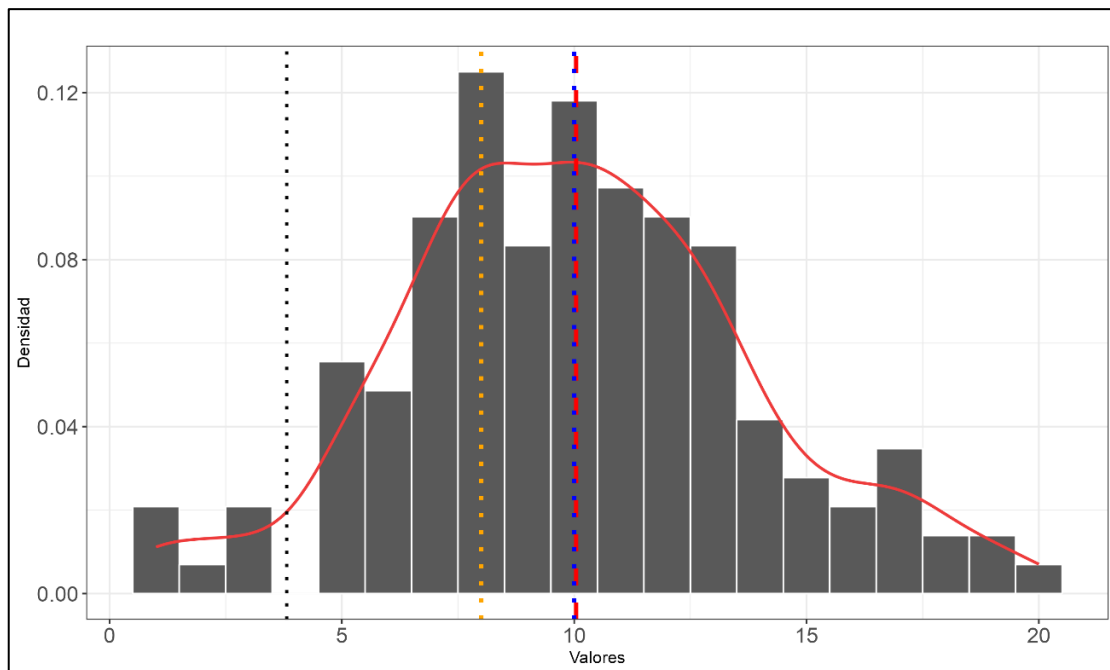
Nota: Estadísticas descriptivas de la serie feminicidios periodos 2010-2021.

A continuación, en la **Figura 13** se muestra el histograma donde se aprecia la distribución de los datos, los datos reportan, un promedio de casos de feminicidio del  $\bar{x} = 10.04$  (línea roja), seguido de una mediana con un número de casos de  $\tilde{x} = 10$  (línea azul) y el número de casos que más veces se repite está denominada como  $mode = 8$  (línea amarilla), además los datos se distribuyen con una desviación estándar de  $\sigma = 3.81$  (línea verde) con respecto a su promedio, es decir que tiene una desviación en porcentaje del 37.97%.

Con la información del párrafo anterior afirmamos que, las distribuciones de los datos fluctúan entorno a su promedio, mediana y moda ( $\bar{x} = \tilde{x}$ ), por lo tanto, la distribución de los datos sigue una distribución normal.

### Figura 13

*Histograma de la serie casos de feminicidios 2010-2021*



Nota: Fuente: Elaboración propia del autor.

## PRUEBAS DE HIPOTESIS DE NORMALIDAD

### TEST DE SHAPIRO-WILK

#### 1. Planteamiento de hipótesis

$H_0$  La muestra proviene de una población con una distribución normal.

$H_1$  La muestra no proviene de una población con una distribución normal

#### 2. Regla de decisión

Si:  $p.value < 0.05$  , Rechazamos la hipótesis nula.

$p.value > 0.05$  , No rechazamos la hipótesis nula.

**Tabla 5**

*Test de Shapiro-Wilk*

statistic	p.value	method
1.0	0.1283**	Shapiro-Wilk normality test

**Signif. codes: 0 <= '\*\*\*' < 0.001 <'\*\*' <0.01 '\*' < 0.05**

Nota: Elaboración propia del autor.

#### 3. Decisión

Dado que el  $p.value = 0.1283$  es mayor que el nivel de significancia típico de  $\alpha = 0.05$ , no rechazamos la hipótesis nula. Por lo tanto, hay evidencia suficiente para afirmar que la serie temporal casos de feminicidio en el Perú 2010-2021 proviene de una población con una distribución normal.

## PRUEBA DE ESTACIONARIEDAD

### TEST DE DICKEY FULLER

#### 1. Planteamiento de hipótesis

$H_0$ : Presenta raíz unitaria (serie no estacionario).

$H_\alpha$ : No presenta raíz unitaria (serie estacionaria).

#### 2. Regla de decisión

Si:  $p. value < 0.05$  , Rechazamos la hipótesis nula.

$p. value > 0.05$  , No rechazamos la hipótesis nula.

**Tabla 6**

*Test de Dickey Fuller*

statistic	p.value	parameter	method	alternative
-5.9	0.0100**	5.0	Augmented Dickey-fuller test	stationary

**Signif. codes: 0 <= '\*\*\*\*' < 0.001 < '\*\*\*' < 0.01 '\*' < 0.05**

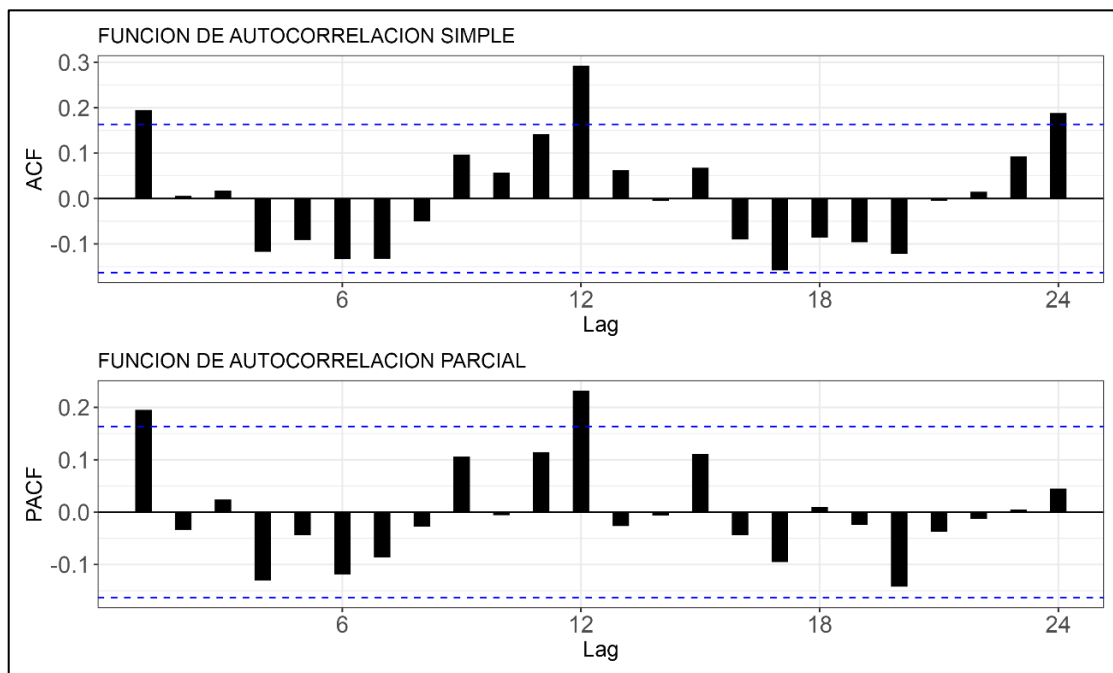
Nota: Elaboración propia del autor.

#### 3. Decisión

Dado que el  $p. value = 0.0100$  es menor que el nivel de significancia típico de  $\alpha = 0.05$ , rechazamos la hipótesis nula. Por lo tanto, hay evidencia suficiente para afirmar que la serie temporal casos de feminicidio en el Perú 2010-2021 es estacionaria y no muestra una raíz unitaria.

**Figura 14**

*Función De Autocorrelación Simple y Parcial de la serie original.*



Nota: Elaboración propia del autor.

Una vez determinado la estacionariedad de la serie casos de feminicidios en el Perú 2010-2021, el siguiente paso es encontrar el modelo *ARIMA* mediante las funciones de autocorrelación simple y parcial tal como lo propone la metodología Box – Jenkins, sin embargo, para la identificación del modelo *ARIMA* emplearemos la función `auto.arima()` del software R versión 4.3.1. , que sirve para identificar modelos *ARIMA* de forma automática.

Al respecto (Donado & Huertas, 2018) sobre la utilización de la función `auto.arima()` en el software r, los investigadores concluyen en su investigación denominada "evaluación crítica a la función `auto.arima()` en la identificación de modelos *ARIMA*" que, la identificación automática de modelos *ARIMA* es un buen aliado para los analistas de información, puesto que reduce en gran medida el tiempo de trabajo que se usa para identificar modelos *ARIMA* de la forma clásica, además señalan que, es

importante conocer el método que estamos aplicando, al igual que la manipulación técnica del software *r* para generar modelos *ARIMA* más confiables.

En la **Figura 14** se muestra las funciones de autocorrelación simple, como también de la autocorrelación parcial en su forma gráfica; En el cual la función de autocorrelación simple nos muestra desfases en los retardos 1, 12 y 24 indicando que estos se encuentran fuera de los límites de confianza, por otro lado, la función de autocorrelación parcial nos muestra desfases en los retardos 1 y 12 los mismos que se encuentran fuera de los límites de confianza.

En la **Tabla 7** se detalla los posibles modelos *ARIMA* para la serie casos de feminicidio 2010-2021 que mejor se ajusta a la serie original, la función `auto.arima()` nos plantea 27 modelos *ARIMA*, de los cuales hemos seleccionado 6 modelos para nuestro análisis.

De los 6/27 modelos posibles generados con la función `auto.arima()`, el modelo 1 ***ARIMA*(0, 0, 1) × (1, 0, 1)<sub>12</sub>** cumple con uno de los principales requisitos exigidos por la “metodología Box-Jenkins”, obteniendo los menores valores posibles entre los 5 modelos, en los criterios de información de Akaike(**AIC**) y de Bayes(**BIC**), demostrando de esta forma que el modelo es óptimo para representar a la serie casos de feminicidio.

**Tabla 7***Posibles modelos ARIMA*

Nº	MODELOS $ARIMA(p, d, q) \times (P, D, Q)$	AIC	BIC
1	$ARIMA(0, 0, 1) \times (1, 0, 1)_{12}$	781.39	796.24
2	$ARIMA(1, 0, 0) \times (1, 0, 1)_{12}$	781.50	796.34
3	$ARIMA(0, 0, 0) \times (1, 0, 1)_{12}$	782.55	794.43
4	$ARIMA(0, 0, 1) \times (1, 0, 0)_{12}$	782.70	794.58
5	$ARIMA(1, 0, 0) \times (1, 0, 0)_{12}$	782.80	794.67
6	$ARIMA(0, 0, 1) \times (2, 0, 0)_{12}$	782.90	797.75
⋮	⋮	⋮	⋮
27	$ARIMA(p, d, q) \times (P, D, Q)$	Inf	Inf

Nota. Modelos ARIMA estacionales. Elaboración propia del autor.

#### 4.2. ESTIMACIÓN DEL MODELO ARIMA

En la **Tabla 8** se muestran los coeficientes estimados del modelo multiplicativo estacional  $ARIMA(0, 0, 1) \times (1, 0, 1)_{12}$ . El término autorregresivo estacional  $sar1$  tiene un coeficiente de  $\Phi_1 = 0.77771$ , lo que indica que hay una relación positiva significativa entre las observaciones pasadas y la actual, con un valor  $p.value = 0.05$  extremadamente pequeño ( $0.05 > 0.0001$ ). El término de intercepción (intercept) tiene un valor estimado de  $C = 10.16817$ , lo que sugiere que este es el valor esperado de la serie temporal cuando todos los demás términos del modelo son cero, con un valor  $p.value$  aún más pequeño ( $0.05 > 0.000000000000000022$ ), lo que indica una alta significancia. El término de media móvil estacional ( $sma1$ ) también es significativo con un coeficiente de  $\Theta_1 = 0.54948$ , indicando una relación negativa significativa entre los errores pasados y el error actual, con un valor de significancia de  $0.05 > 0.02885$ ),



mientras que el término de media móvil regular ( $ma_1$ ) del modelo tiene un coeficiente de  $\theta_1 = 0.15179$ , que no muestra significancia a un nivel de confianza típico del 5% ( $0.05 < 0.07094$ ).

En resumen, estos resultados indican que el componente autorregresivo estacional y el término de intercepción son aspectos críticos en este modelo, mientras que el componente de media móvil estacional también tiene importancia, pero menos fuerte, y el componente de media móvil regular no parece ser significativo.

### Tabla 8

*Estimación de parámetros del modelo  $ARIMA(0,0,1) \times (1,0,1)_{12}$*

Parámetros	Estimate	Std. Error	z value	p.value								
ma1	0.15179	0.08405	1.8059	0.07094	.							
sar1	0.77771	0.1816	4.2825	0.00001848	***							
sma1	-0.54948	0.25143	-2.1855	0.02885	*							
Intercept	10.16817	0.58073	17.5094	<0.000000000000000022	***							
Signif. Codes	0	****	0.001	***	0.01	**	0.05	.	0.1	'	'	1

Nota: Elaboración propia del autor.

Una vez culminado con la estimación de los parámetros del modelo multiplicativo estacional ARIMA, procedemos a realizar el cálculo de la ecuación general que se detalla a continuación:

- **Ecuación general del modelo  $ARIMA(0, 0, 1) \times (1, 0, 1)_{12}$**

$$Y_t = c + \theta_1 \varepsilon_{t-1} + \Theta_1 \varepsilon_{t-12} + \Phi_1 \varepsilon_{t-12} + \varepsilon_t$$

Donde

$y_t =$  Es el valor de la serie temporal en el tiempo  $t$ .

$c =$  Es una constante.



$\varepsilon_t =$  Es un término de error aleatorio en el tiempo  $t$ .

$\theta_1 =$  Es el parámetro del componente de media móvil no estacional.

$\Phi_1 =$  Es el parámetro del componente autorregresivo estacional.

$\Theta_1 =$  Es el parámetro del componente de media móvil estacional.

Acto seguido reemplazamos los parámetros estimados detallados en la **Tabla 8** en la ecuación general, quedando la ecuación final del modelo univariante de la siguiente forma:

- **Reemplazando los parámetros en la ecuación general**

$$Y_t = 10.16817 + 0.15179\varepsilon_{t-1} + 0.54948\varepsilon_{t-12} + 0.77771\varepsilon_{t-12} + \varepsilon_t$$

Donde:

$Y_t =$  Es el valor de la serie temporal en el tiempo  $t$ .

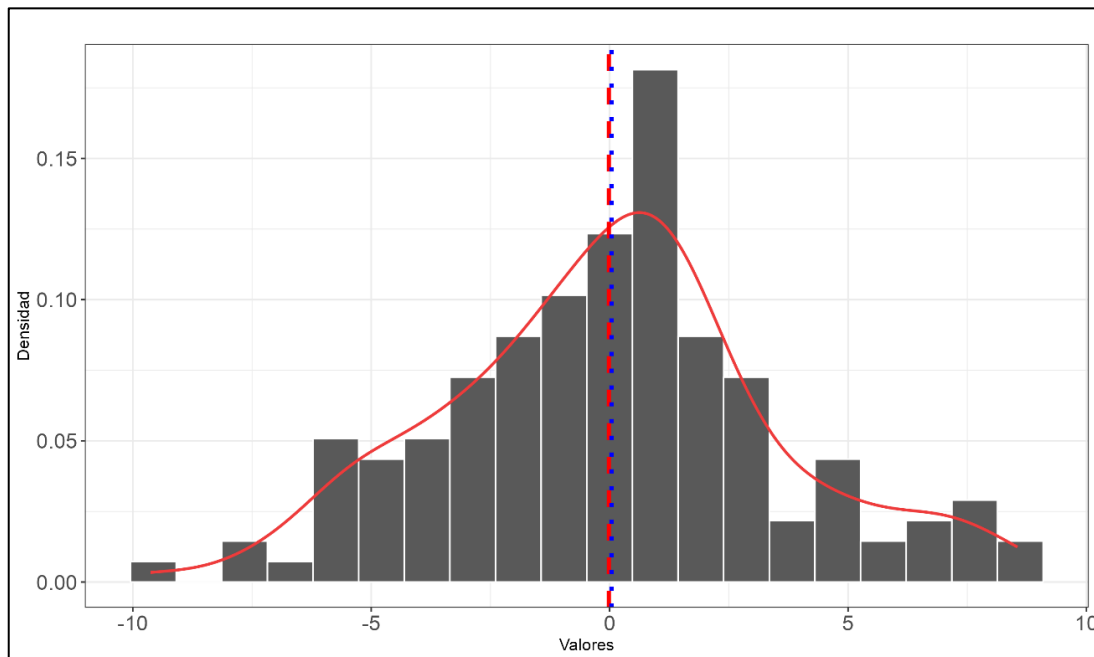
$\varepsilon_t =$  Es un término de error aleatorio en el tiempo  $t$ .

### 4.3. VALIDACIÓN DEL MODELO ARIMA

Para la validación de un modelo ARIMA la “metodología Box-Jenkins” propone dentro de su proceso de análisis, el análisis de residuos a través de una distribución normal, el análisis de residuos mediante las funciones de autocorrelación simple y a su vez de la autocorrelación parcial, y por último análisis de métricas de evaluación y precisión del modelo.

**Figura 15**

*Distribución normal de los residuos del modelo seleccionado*



Nota: Distribución de los residuos del modelo ARIMA seleccionado.

La **Figura 15** muestra la distribución de los residuos del modelo multiplicativo estacional  $ARIMA(0,0,1) \times (1,0,1)_{12}$ , donde se puede apreciar que los residuos del modelo fluctúan entorno a su promedio (línea roja) y mediana (línea azul).

## PRUEBA DE LJUNG-BOX

### 1. Planteamiento de hipótesis

$H_0$ : Los residuos de la serie contiene ruido blanco

$H_1$ : Los residuos de la serie no contienen ruido blanco

### 2. Regla de decisión

Si:  $p.value < 0.05$  , Rechazamos la hipótesis nula.

$p.value > 0.05$  , No rechazamos la hipótesis nula.

**Tabla 9**

*Test de Box-Ljung*

statistic	p.value	parameter	method
0.0	0.9781	1.0	Box-Ljung Test

**Signif. codes: 0 <= '\*\*\*' < 0.001 < '\*\*' < 0.01 < '\*' < 0.05**

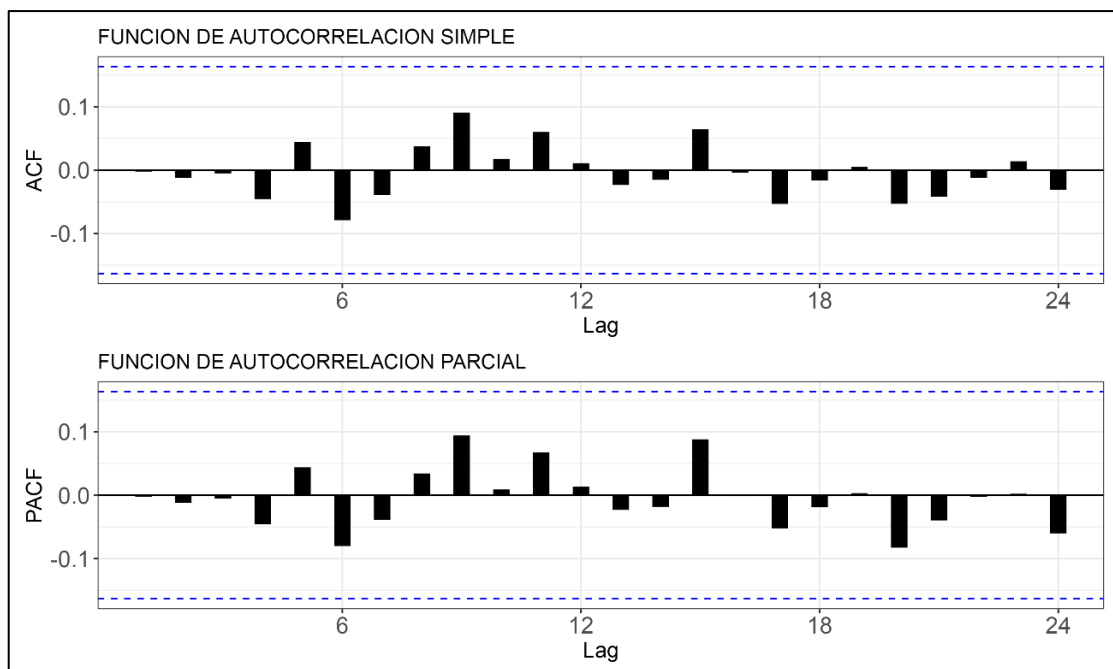
Nota: Test de Box – Ljung para medir la presencia de ruido blanco. Elaboración propia del autor.

### 3. Decisión

Dado que el  $p.value = 0.9781$  es mayor que el nivel de significancia típico de  $\alpha = 0.05$ , no rechazamos la hipótesis nula. Por lo tanto, hay evidencia suficiente para afirmar que, los residuos del modelo  $ARIMA(0,0,1) \times (1,0,1)_{12}$  contiene ruido blanco, es decir que son completamente aleatorios.

**Figura 16**

*Función De Autocorrelación Simple Y Parcial de residuos del modelo seleccionado*



Nota: Autocorrelación simple y parcial de los residuos del modelo.

En la **Figura 16** se muestra la función de autocorrelación simple y parcial de los residuos del modelo  $ARIMA(0,0,1) \times (1,0,1)_{12}$ , donde se puede observar que no hay

ningún desfase en los retardos en ambas funciones de autocorrelación, indicando que no hay ninguna correlación lineal significativa entre los valores.

Por último, con respecto a la evaluación del modelo en cuanto a su precisión, el modelo seleccionado **modelo 1** obtiene valores inferiores en la métricas de medición y precisión frente a los 5 modelos candidato, el modelo que hemos seleccionado tiene un grado de precisión moderada con un valor *MAPE* del 47.36043%, esto quiere decir que por cada pronostico que se haga este estará desviado en 3.498024 unidades, que viene hacer el RMSE del modelo, el MASE es una medida de precisión del modelo con respecto a otro modelo de referencia mientras este sea  $MASE < 1$  el modelo será aún más preciso, en nuestro caso el modelo 1 ha obtenido un  $MASE = 0.775220$  siendo este menor a 1, esto quiere decir que el modelo es preciso frente a los otro modelo de referencia.

**Tabla 10**

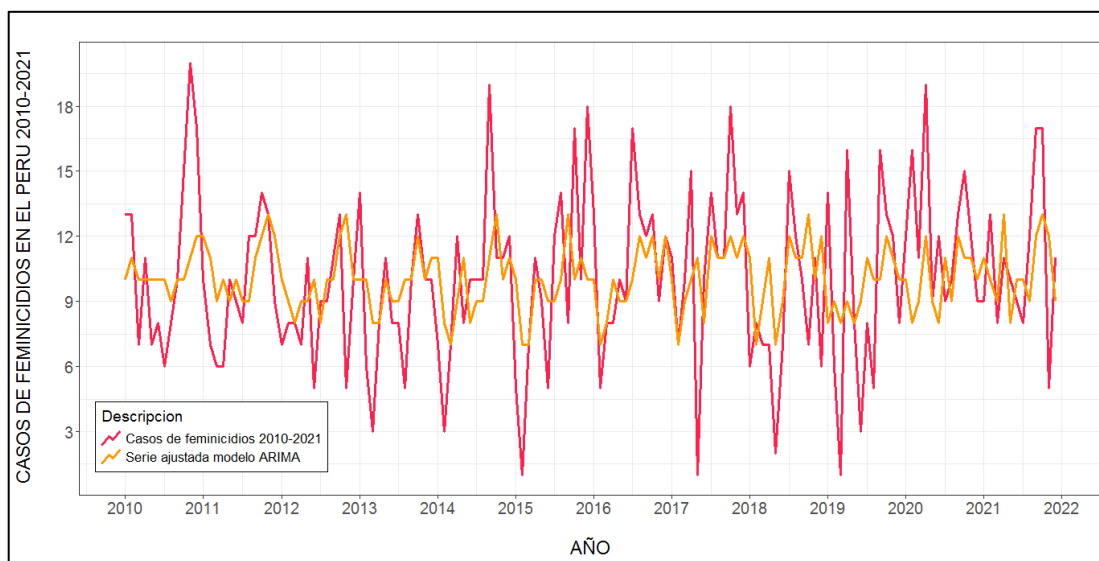
*Métricas de precisión de los modelos ARIMA*

MODELOS Nº	METRICAS DE EVALUACION DEL MODELO				
		RMSE	MAE	MAPE	MASE
1	$ARIMA(0, 0, 1) \times (1, 0, 1)_{12}$	3.498024	2.707398	47.36043	0.775220
2	$ARIMA(1, 0, 0) \times (1, 0, 1)_{12}$	3.499551	2.706659	47.48927	0.775009
3	$ARIMA(0, 0, 0) \times (1, 0, 1)_{12}$	3.532088	2.734285	48.00135	0.782919
4	$ARIMA(0, 0, 1) \times (1, 0, 0)_{12}$	3.549489	2.755621	47.95959	0.789028
5	$ARIMA(1, 0, 0) \times (1, 0, 0)_{12}$	3.550610	2.755565	48.12808	0.789012
6	$ARIMA(0, 0, 1) \times (2, 0, 0)_{12}$	3.523077	2.731862	47.79428	0.782225

Nota: Fuente: Elaboración propia del autor.

**Figura 17**

*Serie original con respecto a la serie ajustada del modelo  $ARIMA(0,0,1) \times (1,0,1)_{12}$*



Nota: Serie original casos de femicidio en el Perú periodo 2010-2021, realizado en el software r v4.3. 1..

Fuente: Elaboración propia de los autores.

En la **Figura 17** se muestra el ajuste del modelo  $ARIMA(0,0,1) \times (1,0,1)_{12}$  con respecto a la serie casos de femicidio en el Perú 2010–2021, los valores generados por el modelo a través de un modelo matemático (línea amarilla) para representar a la serie original (línea roja) se ajustan perfectamente con una desviación en promedio por cada pronóstico de un valor  $MAE = 2.707398 \approx 3$ , es decir de una variación de 3 feminicios por pronostico con respecto al valor real.

#### 4.4. PRONÓSTICO

En la **Tabla 11** se detalla los pronósticos para el año 2022 de la variable casos de feminicios en el Perú 2010-2021, con un nivel de confianza del 95%.

Los pronósticos demuestran que no habrá un incremento ni decrecimiento de casos de feminicios para el año 2022; al contrario, el número de casos de feminicio se mantiene constante con un promedio de 11 casos por mes, el feminicidio es un caso de

estudio bastante amplio que involucra a varias instituciones de un país, en el Perú los casos de feminicidio no son frecuentes debido a que estos hechos se encuentran penados.

**Tabla 11**

*Valores pronosticados casos de feminicidio en el Perú 2022*

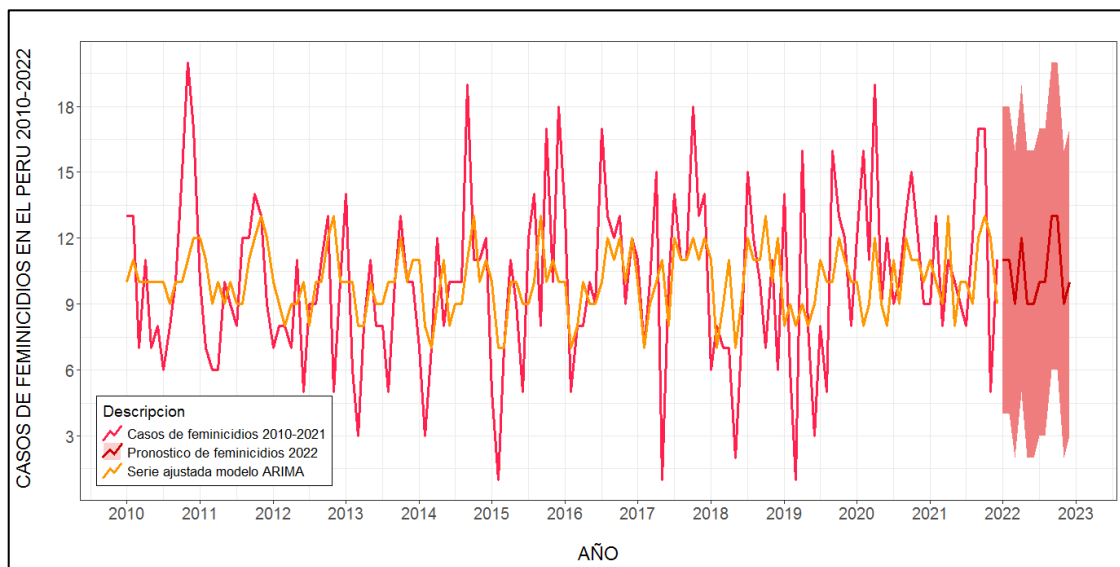
N°	Mes	Año	Pronóstico	Límites de confianza	
				Lo 95	Hi 95
1	Enero	2022	11	4	18
2	Febrero	2022	11	4	18
3	Marzo	2022	9	2	16
4	Abril	2022	12	5	19
5	Mayo	2022	9	2	16
6	Junio	2022	9	2	16
7	Julio	2022	10	3	17
8	Agosto	2022	10	3	17
9	Septiembre	2022	13	6	20
10	Octubre	2022	13	6	20
11	Noviembre	2022	9	2	16
12	Diciembre	2022	10	3	17

Nota: Elaboración propia del autor.

En la **Figura 18** se muestra el periodograma de los pronósticos casos de feminicidios para el año 2022, donde se aprecia gráficamente los pronósticos de feminicidio con un intervalo de confianza del 95% (línea roja sombreada).

## Figura 18

### *Pronostico de casos de feminicidios en el Perú 2022*



Nota: Periodograma de la serie original, serie ajustada y el pronóstico. Elaboración propia del autor.

## 4.5. DISCUSIÓN

Esta investigación abarca el pronóstico de casos de feminicidio en el territorio del Perú para el año 2022, aplicando la “metodología Box- Jenkins” presentada en 1976 que, sirve para analizar una serie temporal e identificar un modelo univariante que se ajuste perfectamente a la serie en estudio y posteriormente realizar pronósticos (Marcelo Barreto & Villareal Escate, 2021).

El objetivo general de esta investigación fue “determinar un modelo univariante que mejor se ajusta a la serie histórica mensual de Feminicidios del Perú, periodo 2010-2021 que tiene como finalidad describir y pronosticar el comportamiento de las variaciones de dicha serie”, para el cumplimiento de este objetivo se aplicaron las 4 fases de la metodología Box-Jenkins (identificación, estimación, validación y pronóstico), llegando a determinar que el mejor modelo univariante que representa a la serie es el



modelo multiplicativo estacional  $ARIMA(0,0,1) \times (1,0,1)_{12}$  representada por la ecuación general

$$Y_t = 10.16817 + 0.15179\varepsilon_{t-1} + 0.54948\varepsilon_{t-12} + 0.77771\varepsilon_{t-12} + \varepsilon_t, \quad \text{el}$$

modelo ha sido seleccionado a través de los criterios de información de Akaike (AIC) y de Bayes (BIC) obteniendo valores inferiores frente a otros 27 modelos candidatos.

$AIC = 781.39$  y  $BIC = 796.24$  lo que indica un equilibrio entre el ajuste del modelo. (Jujulaca Chura, 2019) en su investigación selecciona el mejor modelo univariante siguiendo las 4 fases de la metodología Box-Jenkins, y mediante la selección del mínimo valor en los criterios de información de Akaike (AIC) y de Bayes (BIC) selecciona el modelo  $ARIMA(1,1,0) \times (0,1,1)_{12}$  en comparación a otros modelos. Por otro lado (Acosta Bravo & Chapoñan Bances, 2019) determina el modelo multiplicativo  $ARIMA(1,1,1) \times (2,1,0)_{12}$  empleando la metodología Box – Jenkins, seleccionando el mínimo valor obtenido en los criterios de información de Akaike y de Bayes. En ambas investigaciones se emplea la metodología Box-Jenkins para determinar un modelo ARIMA que mejor se ajuste a la serie histórica, comprobando nuestra hipótesis general.

La metodología *ARIMA* plantea el estricto cumplimiento de cuatro fases para determinar un modelo *ARIMA* (identificar, estimar, validar y pronosticar) de esta forma se asegura que el modelo sea el más óptimo y preciso posible, el modelo que hemos determinado es el modelo  $ARIMA(0,0,1) \times (1,0,1)_{12}$  se seleccionó en bases a los criterios de información de Akaike y de Bayes, en la fase de estimación se estiman los parámetros obteniendo los siguientes coeficientes un  $ma1 = 0.15179$ ,  $sar1 = 0.777771$ ,  $sma1 = -0.54948$  y una constante de  $c = 10.16817$ , los parámetros estimados fueron significativos a un nivel de confianza del 5% como se detalla en la **Tabla 8**, al respecto (Alva Yaipén, 2017b) identifica un modelo  $ARIMA(13,0,13)$  para



pronosticar el número de defunciones en el servicio de cardiología y estima los parámetros en el programa EViews obteniendo estimadores  $MA(13) = -0.906842$   $AR13 = 0.383415$  Y  $c = 0.453957$  con una significancia menor del 5% indicando un buen ajuste, resultados que coinciden con nuestra investigación demostrando nuestra primera hipótesis específica.

Con respecto a nuestra segunda hipótesis específica, nuestra investigación plantea una evolución constante, se debe de tener en cuenta que nuestro modelo  $ARIMA(0,0,1) \times (1,0,1)_{12}$  tiene un grado de precisión del 47% que viene hacer nuestro valor  $MAPE = 47\%$ , nuestro modelo contiene una desviación por cada pronóstico de 3 casos de feminicidio que probablemente podría ocurrir en el año 2022. Al respecto (Acosta Bravo & Chapoñan Bances, 2019) refiere que obtener un valor MAPE menor del 10% indica la eficiencia y precisión de un modelo ARIMA, puesto que en su investigación obtiene un valor  $MAPE = 1.58\%$ , hallazgo que no coincide con nuestra investigación, por otro lado (Alva Yaipén, 2017b) obtiene un valor  $MAPE = 35\%$  lo cual indica un precisión en los pronósticos moderada, hallazgo que concuerda con nuestros resultados, ahora bien, el pronóstico de caso de feminicidio en el Perú para el año 2022 no se incrementa ni decrece, al contrario su evolución es constante a lo largo de los meses del año 2022 con un promedio de 11 casos por meses, esto se debe a que el feminicidio en el Perú y en cualquier parte del Mundo se encuentra penalizado, si bien nuestro modelo tiene una precisión moderada no podemos afirmar que habrá un incremento de casos de feminicidio en el año 2022, esto se puede explicar a diversos factores que influyen en el comportamiento de los seres humanos.

## V. CONCLUSIONES

**PRIMERA:** El mejor modelo univariante que representa un mejor ajuste a la serie histórica mensual de feminicidios en el Perú 2010-2021 está representada por la ecuación:  $Y_t = 10.16817 + 0.15179\varepsilon_{t-1} + 0.54948\varepsilon_{t-12} + 0.77771\varepsilon_{t-12} + \varepsilon_t$ , esta ecuación cumple con todos requisitos que contempla la metodología Box-Jenkins.

**SEGUNDA:** El modelo multiplicativo estacional  $ARIMA(0,0,1) \times (1,0,1)_{12}$ , es el mejor modelo identificado producto del análisis de acuerdo a la metodología Box-Jenkins, presentando valores mínimos para su selección tomando en cuenta los criterios de información de Akaike (AIC) y de Bayes (BIC) obteniendo valores mínimos de  $AIC = 781.39$  y  $BIC = 796.24$ .

**TERCERA:** Los pronósticos alcanzados producto de esta investigación se detallan en la **Tabla 11**, de acuerdo con la información que se presenta, el número de casos de feminicidio para el año 2022 no presentará incrementos ni decrementos, al contrario, se mantendrá constante a lo largo del año 2022 con un promedio de 11 casos por cada mes.



## VI. RECOMENDACIONES

**PRIMERA:** Recomendamos al programa Nacional para la Prevención y Erradicación de la Violencia contra las Mujeres y los Integrantes del Grupo Familiar – AURORA, el uso de esta investigación como referencia en la aplicación y creación de políticas públicas para la prevención de casos de feminicidio en el Perú.

**SEGUNDA:** Al Decano de la Facultad de Ingeniería Estadística e Informática crear convenios interinstitucionales, para desarrollar de manera más compleja investigaciones relacionadas en distintos campos de estudio como salud, educación, minería, industria, energía, entre otros, ello favorecerá en la toma de decisiones basadas en información.

**TERCERA:** Se recomienda para futuras investigaciones en el campo del análisis de feminicidios en el Perú consideren la exploración de modelos de series temporales más complejos, que incorporen no solo variables demográficas y socioeconómicas, sino también factores culturales y políticos que podrían influir en la incidencia de este fenómeno.



## VII. REFERENCIAS BIBLIOGRÁFICAS

- Abril, J. C. (2011). *Análisis de la evolución de las técnicas de series tiempo: un enfoque unificado*.
- Acosta Bravo, R. E., & Chapoñan Bances, D. Y. (2019). Modelo De Series De Tiempo Para Describir La Producción Mensual De Espárrago En El Perú, Periodo Enero 2007 – Diciembre 2018. *Universidad Nacional Pedro Ruiz Gallo*, 113, 92.
- Acosta Bravo, R. E., & Chapoñan Bances, D. Y. (2022). *Modelo de serie de tiempo para describir la producción mensual de espárrago en el Perú, período enero 2007 – diciembre 2018*. <http://repositorio.unprg.edu.pe/handle/20.500.12893/10670>
- Alva Yaipén, K. S. (2017a). Modelo de pronóstico de las defunciones en los servicios de hospitalización del Hospital Belén de Trujillo. *Universidad Nacional de Trujillo*. <http://dspace.unitru.edu.pe/handle/UNITRU/10629>
- Alva Yaipén, K. S. (2017b). *Modelo de pronóstico de las defunciones en los servicios de hospitalización del Hospital Belén de Trujillo*.
- Alvazzi del Frate, A. (2011). *When the victim is a woman. In Global Burden of Armed Violence* 2011.(pp. 113-144). [http://www.genevadeclaration.org/fileadmin/docs/GBAV2/GBAV2011\\_CH4.pdf](http://www.genevadeclaration.org/fileadmin/docs/GBAV2/GBAV2011_CH4.pdf)
- Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock Price Prediction Using the ARIMA Model. *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, 106-112. <https://doi.org/10.1109/UKSim.2014.67>
- AURORA. (2021). *Cifras de Violencia contra las Mujeres*. <https://portalestadistico.aurora.gob.pe/wp-content/uploads/2022/01/Cartilla-Estadistica-AURORA-Diciembre-2021-1.pdf>
- Batanero, C. (2001). *Didáctica de la Estadística*.



- Bernáldez Jaimes, G. B., Rodríguez Aguilar, B., Bernáldez Jaimes, G. B., & Rodríguez Aguilar, B. (2023). Consecuencias psicosociales en familiares víctimas de feminicidio en México. *Acta universitaria*, 33. <https://doi.org/10.15174/au.2023.3586>
- Bernardí Cabrer. (2004). *Econometría empresarial II: Vol. Tema 8*.
- Bodelón, E. (2014). Violencia institucional y violencia de género. *Anales de la Cátedra Francisco Suárez*, 48, 131-155. <https://doi.org/10.30827/acfs.v48i0.2783>
- Caballero Díaz, F. (2011). *Selección de modelos mediante criterios de información en análisis factorial: Aspectos teóricos y computacionales*. Universidad de Granada.
- Carmona Arce, M., & Carrion Rosales, H. (2015). *Potencia de la prueba estadística de normalidad JARQUE-BERA frente a las pruebas de ANDERSON-DARLING, JARQUE-BERA ROBUSTA, CHI-CUADRADA, CHEN-SHAPIRO y SHAPIRO-WILK*. <http://ri.uaemex.mx/handle/20.500.11799/94337>
- Carranza Trujillo, M. R. (2019). Análisis estadístico de las exportaciones de bienes en la Comunidad Andina, Enero 2010- Junio 2018. *Universidad Nacional de Trujillo*. <http://dspace.unitru.edu.pe/handle/UNITRU/15104>
- Carrigan, M. (s. f.). *Femicide Legislation: Lessons from Latin America*.
- Chatfield, C. (1987). *The analysis of time series an introduction*. Chapman & Hall.
- Chinlli, C. M. (2021). *Modelización de Series Temporales modelos clásicos y SARIMA*. Universidad de Granada.
- Contento Rubio, M. R. (2019). Estadística con aplicaciones en R. En *Instname:Universidad de Bogotá Jorge Tadeo Lozano*. Universidad de Bogotá Jorge Tadeo Lozano. <https://doi.org/10/21660>



- Coutin Marie, G. (2007). Utilización de modelos ARIMA para la vigilancia de enfermedades transmisibles. *Revista Cubana de Salud Pública*, 33(2), 0-0.
- Cryer, J. D., & Chan, K. (2008). *Time Series Analysis With Applications in R* (Second Edition). Springer.
- Damián Llatas, M. R., & Sandoval Santamaría, N. J. (2018). *Modelo óptimo de Pronóstico del índice mensual de Producción de Electricidad"-Perú en el Periodo 2006 – 2015*. <http://repositorio.unprg.edu.pe/handle/20.500.12893/2192>
- Donado, S., & Huertas, N. (2018). *Evaluación Crítica De La Función auto.arima De R En La Identificación De Procesos De Series De Tiempo* (Número 201326730).
- Fernández Casal, R., Costa Bouzas, J., & Oviedo de la Fuente, M. (2021, Setiembre). *Aprendizaje Estadístico*.
- Gobierno del Perú. (2015, noviembre 23). *Ley 30364—Ley para prevenir, sancionar y erradicar la violencia contra las mujeres y los integrantes del grupo familiar*. gov.pe. <https://www.gob.pe/institucion/congreso-de-la-republica/normas-legales/3212371-30364>
- González Casimiro, M. P. (2009). *Análisis de series temporales: Modelos ARIMA*. Economía Aplicada III/Ekonomia Aplikatua III, UPV/EHU. <https://doi.org/10/12492>
- Guerra Rodríguez, J. (2022). *Fundamentos y variantes de los modelos ARIMA para el análisis de series temporales: Aplicación a la estadística universitaria*. <https://riull.ull.es/xmlui/handle/915/28456>
- Hankhe, J. E., & Reitsch, A. G. (1996). *Pronosticos en los negocios* (Quinta Edicion). Prentice Hall Inc.
- Heise, L., & Garcia Moreno, C. (s. f.). *Violence by Intimate Partners (From World Report on Violence and Health, P 87-121, 2002, Etienne G. Krug, Linda L. Dahlberg, et*



al., eds. —See NCJ-197425) | *Office of Justice Programs*. Recuperado 8 de febrero de 2023, de <https://www.ojp.gov/ncjrs/virtual-library/abstracts/violence-intimate-partners-world-report-violence-and-health-p-87>

Jiménez Guerrero, J. F., Sánchez Fernández, R., & Gázquez Abad, J. C. (2006). La capacidad predictiva en los métodos Box-Jenkins y Holt-Winters: Una aplicación al sector turístico. *Revista europea de dirección y economía de la empresa*, 15(3), 185-198.

J.M. CARIDAD, & OCERIN. (2005). *Modelos econométricos multiecuacionales: Predicción económica y series temporales*. Editorial Reverté S.A.

Juculaca Chura, J. (2019). Modelo univariante para predecir el número de casos de infecciones respiratorias agudas, neumonía y defunciones en niños menores de 5 años en la Dirección Regional de Salud Puno – 2018. *Universidad Nacional del Altiplano*. <https://repositorio.unap.edu.pe/handle/20.500.14082/10699>

Jujulaca Chura, J. (2019). *Modelo univariante para predecir el número de casos de infecciones respiratorias agudas, neumonía y defunciones en niños menores de 5 años en la Dirección Regional de Salud Puno–2018*. UNIVERSIDAD NACIONAL DEL ALTIPLANO.

*Ley que modifica el Código Penal y el Código de los Niños y Adolescentes—LEY - N° 30819—CONGRESO DE LA REPUBLICA*. (2018, julio 13). <https://busquedas.elperuano.pe/dispositivo/NL/undefined/dispositivo/NL/1669642-1>

López, L. I. M., & Rissi, P. S. (2008, julio). *Informe Nacional sobre Femicidio en Perú*. SINCO editores SAC.

López Tito, J. A. (2018). Modelos univariantes para el pronóstico de número de casos de aborto y preeclampsia de hospitalización del servicio de obstetricia del Hospital





- Regional Manuel Núñez Butrón de Puno periodo 2002 – 2018-II. *Universidad Nacional del Altiplano*.  
<https://repositorio.unap.edu.pe/handle/20.500.14082/9630>
- Machaca Quispe, C. U. (2017). Modelo univariante para la predicción de la desnutrición crónica de los niños menores de cinco años en el Hospital Regional Manuel Núñez Butrón, Puno 2012—2016. *Universidad Nacional del Altiplano*.  
<https://repositorio.unap.edu.pe/handle/20.500.14082/9386>
- Mamani Condori, B. (2020). Modelo para la predicción de la demanda mensual de horas de maquinaria agrícola en la municipalidad distrital de Caracoto—2020. *Universidad Nacional del Altiplano*.  
<https://repositorio.unap.edu.pe/handle/20.500.14082/14177>
- Marcelo Barreto, E. A., & Villareal Escate, L. G. (2021). *El Modelo Estocástico Univariante ARIMA Como Herramienta Predictiva De La Demanda De Energía Eléctrica Residencial Del Sistema Eléctrico Cusco*. Universidad Nacional de Ingeniería.
- Matilla García, M., Pérez Pascual, P., & Sanz Carnero, B. (2017). *ECONOMETRÍA Y PREDICCIÓN* (Segunda Edición). McGraw-Hill.
- Melo Mayta, E. (2016). Modelo de predicción mensual de mortalidad general intrahospitalaria en el Hospital Regional Manuel Núñez Butrón-Puno, 2008-2016-I. *Universidad Nacional del Altiplano*.  
<https://repositorio.unap.edu.pe/handle/20.500.14082/2092>
- Ministerio de Salud del Perú. (2009, mayo). *Mortalidad General en el Perú 2001—2006*. SINCO editores SAC.
- Mondal, P., Shit, L., & Goswami, S. (2014). Study of Effectiveness of Time Series Modeling (Arima) in Forecasting Stock Prices. *International Journal of Computer*



*Science, Engineering and Applications*, 4, 13-29.

<https://doi.org/10.5121/ijcsea.2014.4202>

Montenegro García, Á. (2011). *Análisis de series de tiempo* (Primera Edición). Pontificia Universidad Javeriana.

Moreno Pérez, S., & Lizárraga Morales, K. (2017). *Carpeta Informativa. Femicidios: Causas, consecuencias y tendencias*. Estados Unidos Mexicanos - Cámara de Diputados. <http://www5.diputados.gob.mx/index.php/camara/Centros-de-Estudio/CESOP/Novedades/Carpeta-Informativa.-Femicidios-causas-consecuencias-y-tendencias>

OMS. (2005). *Addressing violence against women and achieving the Millennium Development Goals*. World Health Organization. <https://apps.who.int/iris/bitstream/handle/10665/43361/?sequence=1>

OMS. (2024, junio 14). *Prevención de la violencia—OPS/OMS | Organización Panamericana de la Salud*. <https://www.paho.org/es/temas/prevencion-violencia>

ONU. (1993, diciembre 20). *Declaración sobre la eliminación de la violencia contra la mujer*. OHCHR. <https://www.ohchr.org/es/instruments-mechanisms/instruments/declaration-elimination-violence-against-women>

PATH, MRC, WHO, & Intercambios. (2008). *Strengthening Understanding of Femicide: Using Research to Galvanize Action and Accountability*. [https://media.path.org/documents/GVR\\_femicide\\_rpt.pdf](https://media.path.org/documents/GVR_femicide_rpt.pdf)

Peña, D. (2005). *ANÁLISIS DE SERIES TEMPORALES*. Alianza Editorial.

Pérez López, C. (2008). *Técnicas de Análisis Multivariante de Datos: Aplicaciones con SPSS*. PEARSON PRENTICE HALL.



- RAE. (s. f.). *Definición de femicidio—Diccionario panhispánico del español jurídico—*  
RAE. Diccionario panhispánico del español jurídico - Real Academia Española.  
Recuperado 19 de marzo de 2023, de <http://dpej.rae.es/lema/femicidio>
- Ramos Inga, A. P. (2020). Pronóstico de los ingresos tributarios mensuales del gobierno central peruano aplicando redes neuronales y modelos Sarima, en base a los años 2003 – 2018. *Universidad Nacional de Piura*.  
<http://repositorio.unp.edu.pe/handle/20.500.12676/2264>
- Russell, D., & Harmes, R. (2006). *Femicidio: Una perspectiva global*.
- Sanchez Perez, A. L. (2021). *ANÁLISIS DE LA MORTALIDAD A PARTIR DE LA PERSPECTIVA DE GÉNERO, 2018*.  
<http://ri.uaemex.mx/handle/20.500.11799/111650>
- Santiago de la Fuente, F. (2012). *SERIES TEMPORALES, MODELO ARIMA: METODOLOGÍA DE BOX - JENKINS*.
- Soto, G. G., & Cenitagoya Garín, V. (2017). *FEMINICIDIO Y SUICIDIO DE MUJERES POR RAZONES DE GÉNERO: Desafíos y aprendizajes en la Cooperación Sur-Sur en América Latina y el Caribe*. FLACSO-Chile. [www.flacsochile.org](http://www.flacsochile.org)
- Uriel, E. (1985). *Análisis de Series de Tiempo Modelos ARIMA*.
- Villareal, F. (2016, Setiembre). *Introducción a los modelos de Pronósticos*.
- Villegas Zamora, D. A. (2019). La importancia de la estadística aplicada para la toma de decisiones en Marketing. *Revista Investigación y Negocios*, 12(20), 31-44.
- Viviano Llave, T. (2007, marzo). *¿Qué son los centros Emergencia Mujer? Situación actual y perspectivas dentro del proceso de descentralización*.
- WORLD BANK GROUP. (2021). *Combating Gender-Based Violence in Peru: Increasing Awareness and Resources to Prevent Violence Against Women*. World Bank. <https://www.worldbank.org/en/results/2021/04/13/combating-gender->



based-violence-in-peru-increasing-awareness-and-resources-to-prevent-  
violence-against-women

World Health Organization. (2013). *Global and regional estimates of violence against women: Prevalence and health effects of intimate partner violence and non-partner sexual violence.* World Health Organization.  
<https://apps.who.int/iris/handle/10665/85239>



## ANEXOS



## ANEXO 1. Código fuente del proceso de la metodología Box – Jenkins

```
# LIMPIAR MEMORIA
rm(list = ls())
ls()
options(scipen = 999)
options(encoding = "UTF-8")
library(readxl)
library(tidyverse)
library(ggplot2)
library(forecast)
library(tseries)

###EXPORTAMOS LA BASE DE DATOS
femicidio <- read_excel("E:/TRABJOS DE TESIS/TESIS FINESI - FEMENICIDIO/scrip-
R/femicidios/femicidios.xlsx")
View(femicidio)

#### LIMPIEZA DE DATOS (FORMATO DE ANCHO A LARGO)
femicidio_a <- femicidio %>% pivot_longer(cols = -MESES,
                                         names_to = "year",
                                         values_to = "casos") %>%
  mutate( year = seq(as.Date("2010-01-01"),
as.Date("2021-12-31"), by = "month") ) %>%
  arrange(year)
View(femicidio_a)

#TRANSFORMAR LA DATA A UN FORMATO TS
femicidio_ts <- ts(femicidio_a$casos, start = c(2010,1), end = c(2021,12),
frequency = 12 )
print(femicidio_ts)

#### GRAFICAMOS LA DATA
grafico_fem <- ggplot(femicidio_a, aes(x= year,y=casos, color = "Femicidios"))+
  geom_line(size = 1.3,key_glyph = "timeseries")+
  theme_bw()

grafico_fem

#####DESCOMPOSICION DE LA SERIE ORIGINAL FEMENICIDIO
desc_femin<- decompose(femicidio_ts, type='multiplicative')
grafico_desc <- autoplot(desc_femin, lwd = 3)+
  theme_bw()+
  labs(x = "",
       y = "",
       title = " ",
       fill = "")

grafico_desc

# CALCULAR Y GRAFICAR LAS FUNCIONES DE AUTOCORRELACIÓN
acf_a <- ggAcf(femicidio_ts, lag = , lwd = 5) +
  theme_bw()+
  theme( axis.text = element_text(size = 15),
        axis.title = element_text(size = 14L))+
  ggtitle("FUNCION DE AUTOCORRELACION SIMPLE")

pacf_a<- ggPacf(femicidio_ts , lag = , lwd = 5) +
  theme_bw()+
  theme(
    axis.text = element_text(size = 15),
```



```
axis.title = element_text(size = 14L) )+
ggtitle("FUNCION DE AUTOCORRELACION PARCIAL")

funcion <- acf_a/pacf_a

####CURVA NORMAL DE LOS DATOS
graf_normal <-ggplot(feminicidio_a, aes(x = casos)) +
  geom_histogram(aes(y = ..density..), bins = 20, color = "white")+
  geom_density(color = "brown2", alpha = 0.6, size = 0.9) +
  geom_vline(xintercept = media, linetype = "dashed", color = "red", size = 1.3) +
  geom_vline(xintercept = mediana, linetype = "dotted", color = "blue", size = 1.3)
+
  geom_vline(xintercept = desviacion_estandar, linetype = "dotted", color = "black",
size = 1) +
  geom_vline(xintercept = moda, linetype = "dotted", color = "orange", size = 1.3) +
  theme_bw()

##PRUEBA DE NORMALIDAD
feminicidio_test <- ks.test(x = femem_diff$x,"pnorm", mean(femin_diff$x),

##PRUEBA DE DIKEY FULLER AUMENTADO - ADF
shap_femen <- shapiro.test(feminicidio_ts)

##PRUEBA DE DICKEY FULLER AUMENTADO ADF
adf_femen <- adf.test(feminicidio_ts)

#####MODELOS ARIMA
arima1<- Arima(feminicidio_ts, order=c(0,0,1), seasonal=list(order=c(1,0,1),period=12))
arima2<- Arima(feminicidio_ts, order=c(1,0,0), seasonal=list(order=c(1,0,1),period=12))
arima3<- Arima(feminicidio_ts, order=c(0,0,0), seasonal=list(order=c(1,0,1),period=12))
arima4<- Arima(feminicidio_ts, order=c(0,0,1), seasonal=list(order=c(1,0,0),period=12))
arima5<- Arima(feminicidio_ts, order=c(1,0,0), seasonal=list(order=c(1,0,0),period=12))
arima6<- Arima(feminicidio_ts, order=c(0,0,1), seasonal=list(order=c(2,0,0),period=12))

#####BIC Y AIC DEL MODELO ARIMA
# Crear un vector con los nombres de los modelos
modelos <- c("arima1", "arima2", "arima3", "arima4", "arima5", "arima6")

# CALCULAR Y GRAFICAR EL CORRELOGRAMA DE AUTOCORRELACIÓN
acf_b <- ggAcf(arima1$residuals, lag = , lwd = 5) +
  theme_bw()+
  theme( axis.text = element_text(size = 15),
axis.title = element_text(size = 14L))+
  ggtitle("FUNCION DE AUTOCORRELACION SIMPLE")

pacf_b<- ggPacf(arima1$residuals, lag = , lwd = 5) +
  theme_bw()+
  theme(
axis.text = element_text(size = 15),
axis.title = element_text(size = 14L) )+
  ggtitle("FUNCION DE AUTOCORRELACION PARCIAL")

funcion_b <- acf_b/pacf_b

##PRUEBA DE DIKEY FULLER AUMENTADO - ADF
shap_femin_b <- shapiro.test(arima1$residuals)

##PRUEBA DE DICKEY FULLER AUMENTADO ADF
adf_femin_b <- adf.test(arima1$residuals)

#####PRUEBA DE RUIDO BLANCO - LJUNG-BOX
test_ljb <- Box.test(arima1$residuals, type = "Ljung-Box")
```



```
#####ESTIMACION DEL MODELO
estimacion <- coeftest(arima1)

#IMPRIMIR RESULTADOS
predicciones <- forecast(arima1,level = c(95), h = 12)
print(predicciones)

#####AJUSTE DEL MODELO ARIMA
grafico_pronostico <- autoplot(femicidio_ts, series = "Casos de femicidio 2010-
2021", lwd = 1.3, lty = 1, key_glyph = "timeseries")+
  autolayer(fitted(predicciones),series = "Serie ajustada modelo ARIMA", lwd = 1.3,
key_glyph = "timeseries")+
  autolayer(predicciones, series = "Pronostico de femicidios 2022", lwd = 1.3,
lty = 1, size = 1.3)+
  theme_bw()+
grafico_pronostico

#####GRAFICO DE NORMALIDAD
graf_normal_b <- ggplot(residuos, aes(x = x)) +
  geom_histogram(aes(y = ..density..), bins = 20, color = "white")+
  geom_density(color = "brown2", alpha = 0.6, size = 0.9) + geom_vline(xintercept =
mean(residuos$x), linetype = "dashed", color = "red", size = 1.3) +
  geom_vline(xintercept = median(residuos$x), linetype = "dotted", color = "blue",
size = 1.3)
  geom_vline(xintercept = , linetype = "dotted", color = "black", size = 1)
#geom_vline(xintercept = moda, linetype = "dotted", color = "orange", size = 1.3)
theme_bw()
graf_normal_b

#####GRAFICO DE SERIES DE TIEMPO
grafico_pronostico <- autoplot(femicidio_ts, series = "Casos de femicidio 2010-
2021", lwd = 1.3, lty = 1, key_glyph = "timeseries")+
  autolayer(fitted(predicciones),series = "Serie ajustada modelo ARIMA", lwd = 1.3,
key_glyph = "timeseries")+
  autolayer(predicciones, series = "Pronostico de femicidios 2022", lwd = 1.3, lty
= 1, size = 1.3)+
  theme_bw()+
  theme(
    plot.title = element_text(size = 18L,
                              face = "bold",
                              hjust = 0.5),
    axis.title.y = element_text(size = 17L),
    axis.title.x = element_text(size = 17L),
    plot.caption = element_text(size = 14L)
  )
grafico_pronostico
```





## DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo **GABRIELA YAMILET BALCONA CCAMA**,

identificado con DNI **74058951** en mi condición de egresado de:

**Escuela Profesional**,  **Programa de Segunda Especialidad**,  **Programa de Maestría o Doctorado**  
**INGENIERIA ESTADÍSTICA E INFORMÁTICA**

informo que he elaborado el/la  **Tesis** o  **Trabajo de Investigación** denominada:

**“MODELO UNIVARIANTE PARA DESCRIBIR Y PRONOSTICAR EL COMPORTAMIENTO DE LA SERIE HISTÓRICA DE FEMINICIDIOS DEL PERÚ, PERIODO 2010 - 2021”**

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 11 de julio del 2024

FIRMA (obligatoria)



Huella



## AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo **GABRIELA YAMILET BALCONA CCAMA**,  
identificado con DNI **74058951** en mi condición de egresado de:

**Escuela Profesional**,  **Programa de Segunda Especialidad**,  **Programa de Maestría o Doctorado**

**INGENIERIA ESTADÍSTICA E INFORMÁTICA**

,informo que he elaborado el/la  **Tesis** o  **Trabajo de Investigación** denominada:

**“MODELO UNIVARIANTE PARA DESCRIBIR Y PRONOSTICAR EL COMPORTAMIENTO DE LA SERIE HISTÓRICA DE FEMINICIDIOS DEL PERÚ, PERIODO 2010 - 2021”**

para la obtención de  **Grado**,  **Título Profesional** o  **Segunda Especialidad**.

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los “Contenidos”) que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

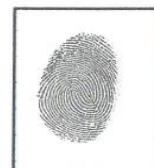
Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

En señal de conformidad, suscribo el presente documento.

Puno 11 de julio del 2024

FIRMA (obligatoria)



Huella



## DECLARACIÓN JURADA DE AUTENTICIDAD DE TESIS

Por el presente documento, Yo **LEIDY MADU LOPEZ MAMANI**,

identificado con DNI **75397192** en mi condición de egresado de:

**Escuela Profesional**,  **Programa de Segunda Especialidad**,  **Programa de Maestría o Doctorado**  
**INGENIERIA ESTADÍSTICA E INFORMÁTICA**

informo que he elaborado el/la  **Tesis** o  **Trabajo de Investigación** denominada:

**“MODELO UNIVARIANTE PARA DESCRIBIR Y PRONOSTICAR EL COMPORTAMIENTO DE LA SERIE HISTÓRICA DE FEMINICIDIOS DEL PERÚ, PERIODO 2010 - 2021”**

Es un tema original.

Declaro que el presente trabajo de tesis es elaborado por mi persona y **no existe plagio/copia** de ninguna naturaleza, en especial de otro documento de investigación (tesis, revista, texto, congreso, o similar) presentado por persona natural o jurídica alguna ante instituciones académicas, profesionales, de investigación o similares, en el país o en el extranjero.

Dejo constancia que las citas de otros autores han sido debidamente identificadas en el trabajo de investigación, por lo que no asumiré como tuyas las opiniones vertidas por terceros, ya sea de fuentes encontradas en medios escritos, digitales o Internet.

Asimismo, ratifico que soy plenamente consciente de todo el contenido de la tesis y asumo la responsabilidad de cualquier error u omisión en el documento, así como de las connotaciones éticas y legales involucradas.

En caso de incumplimiento de esta declaración, me someto a las disposiciones legales vigentes y a las sanciones correspondientes de igual forma me someto a las sanciones establecidas en las Directivas y otras normas internas, así como las que me alcancen del Código Civil y Normas Legales conexas por el incumplimiento del presente compromiso

Puno 11 de julio del 2024

FIRMA (obligatoria)



Huella





## AUTORIZACIÓN PARA EL DEPÓSITO DE TESIS O TRABAJO DE INVESTIGACIÓN EN EL REPOSITORIO INSTITUCIONAL

Por el presente documento, Yo **LEIDY MADU LOPEZ MAMANI**,

identificado con DNI **75397192** en mi condición de egresado de:

**Escuela Profesional**,  **Programa de Segunda Especialidad**,  **Programa de Maestría o Doctorado**

**INGENIERIA ESTADÍSTICA E INFORMÁTICA**

,informo que he elaborado el/la  **Tesis** o  **Trabajo de Investigación** denominada:

**“MODELO UNIVARIANTE PARA DESCRIBIR Y PRONOSTICAR EL COMPORTAMIENTO DE LA SERIE HISTÓRICA DE FEMINICIDIOS DEL PERÚ, PERIODO 2010 - 2021”**

para la obtención de  **Grado**,  **Título Profesional** o  **Segunda Especialidad**.

Por medio del presente documento, afirmo y garantizo ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual sobre los documentos arriba mencionados, las obras, los contenidos, los productos y/o las creaciones en general (en adelante, los “Contenidos”) que serán incluidos en el repositorio institucional de la Universidad Nacional del Altiplano de Puno.

También, doy seguridad de que los contenidos entregados se encuentran libres de toda contraseña, restricción o medida tecnológica de protección, con la finalidad de permitir que se puedan leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna.

Autorizo a la Universidad Nacional del Altiplano de Puno a publicar los Contenidos en el Repositorio Institucional y, en consecuencia, en el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, sobre la base de lo establecido en la Ley N° 30035, sus normas reglamentarias, modificatorias, sustitutorias y conexas, y de acuerdo con las políticas de acceso abierto que la Universidad aplique en relación con sus Repositorios Institucionales. Autorizo expresamente toda consulta y uso de los Contenidos, por parte de cualquier persona, por el tiempo de duración de los derechos patrimoniales de autor y derechos conexos, a título gratuito y a nivel mundial.

En consecuencia, la Universidad tendrá la posibilidad de divulgar y difundir los Contenidos, de manera total o parcial, sin limitación alguna y sin derecho a pago de contraprestación, remuneración ni regalía alguna a favor mío; en los medios, canales y plataformas que la Universidad y/o el Estado de la República del Perú determinen, a nivel mundial, sin restricción geográfica alguna y de manera indefinida, pudiendo crear y/o extraer los metadatos sobre los Contenidos, e incluir los Contenidos en los índices y buscadores que estimen necesarios para promover su difusión.

Autorizo que los Contenidos sean puestos a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

En señal de conformidad, suscribo el presente documento.

Puno 11 de julio del 2024

FIRMA (obligatoria)



Huella