

**UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO**

**FACULTAD DE CIENCIAS FISICAS Y MATEMATICAS**

**ESCUELA PROFESIONAL DE ESTADISTICA**



**TESIS**

Para Optar el Título Profesional de Licenciado en Estadística

**TÍTULO**

**“Modelo de serie de tiempo para pronosticar el número de pacientes con  
cáncer, Hospital Nacional Almanzor Aguinaga Asenjo,**

**Periodo 2006 - 2017”**

Investigador:

Bach. Percy Antony Quispe Chanta

Asesor:

Dr. Alfonso Tesen Arroyo

Lambayeque, enero 2021

# UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO

## FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS ESCUELA PROFESIONAL DE ESTADÍSTICA




---

M.Sc. Débora Esther Mejía Pacheco  
Presidenta



---

Dr. Wilver Omero Rodríguez López  
Secretario



---

M.Sc. Grimaldo Dermalí Benavides Campos  
Vocal



---

Dr. Alfonso Tesen Arroyo  
Asesor



## **ACTA DE SUSTENTACIÓN VIRTUAL N° 007-2021-D/FACFyM**

Siendo las 9:00 am del día jueves 11 de febrero del 2021, se reunieron vía plataforma virtual, <https://meet.google.com/khx-vskn-kiy>, los miembros del jurado evaluador de la Tesis titulada: **Modelo de serie de tiempo para pronosticar el número de pacientes con cáncer, Hospital Nacional Almanzor Aguinaga Asenjo, Periodo 2006 – 2017.**

Designados por Resolución N° 1225-2017-D/FACFyM de fecha 15 de diciembre 2017

Con la finalidad de evaluar y calificar la sustentación de la tesis antes mencionada, conformada por los siguientes docentes:

<b>M.Sc. Débora Esther Mejía Pacheco</b>	<b>Presidente</b>
<b>Dr. Wilver Omero Rodríguez López</b>	<b>Secretario</b>
<b>M.Sc. Grimaldo Dermalí Benavides Campos</b>	<b>Vocal</b>

La tesis fue asesorada por el Dr. Alfonso Tesén Arroyo, nombrado por Resolución N° N°408-2017-D/FACFyM de fecha 25 de Abril 2017.

El Acto de Sustentación fue autorizado por Resolución N° 080-2021-VIRTUAL-D/FACFyM de fecha 05 de Febrero 2021.

La Tesis fue presentada y sustentada por el Bachiller: Percy Antony Quispe Chanta, y tuvo una duración de 30 minutos.

Después de la sustentación, y absueltas las preguntas y observaciones de los miembros del jurado se procedió a la calificación respectiva, otorgándole el Calificativo de 17(Dieciséte) en la escala vigesimal, mención **Bueno**.

Por lo que queda apto para obtener el Título Profesional de **Licenciado en Estadística**, de acuerdo con la Ley Universitaria 30220 y la normatividad vigente de la Facultad de Ciencias Físicas y Matemáticas y la Universidad Nacional Pedro Ruiz Gallo.

Siendo las 10:20. am se dio por concluido el presente acto académico, dándose conformidad al presente acto con la firma de los miembros del jurado.

**M.Sc. Débora Esther Mejía Pacheco**  
**Presidente**

**Dr. Wilver Omero Rodríguez López**  
**Secretario**

**M.Sc. Grimaldo Dermalí Benavides Campos**  
**Vocal**

**Dr. Alfonso Tesén Arroyo**  
**Asesor**

### **Declaración jurada de Originalidad**

Yo, Percy Antony Quispe Chanta, investigador principal, y Dr. Alfonso Tesen Arroyo, asesor del trabajo de investigación “Modelo de serie de tiempo para pronosticar el número de pacientes con cáncer, Hospital Nacional Almanzor Aguinaga Asenjo. Periodo 2006 – 2017”, declaramos bajo juramento que este trabajo no ha sido plagiado, ni contiene datos falsos. En caso se demostrará lo contrario, asumo responsablemente la anulación de este informe y por ende el proceso administrativo a que hubiera lugar. Que puede conducir a la anulación del título o grado emitido como consecuencia de este informe.

Lambayeque, 05 de enero del 2021



---

Bach. Percy Antony Quispe Chanta  
Investigador



---

Dr. Alfonso Tesen Arroyo  
Asesor

## **Dedicatoria**

Con amor a mis padres, quienes son el pilar fundamental en mi vida ya que, con su amor, apoyo y su comprensión me ayudaron a salir adelante y superar los retos que la vida me puso. A Kenia Katherine, por su apoyo incondicional durante todo este proceso. También con mucho respeto y cariño, a la memoria de la señora Nélida Idelsa Alarcón Díaz.

## **Agradecimiento**

**El autor** Agradezco a Dios, por permitirme la oportunidad de haber logrado cumplir mis metas trazadas.

A mi asesor, por su apoyo constante para el desarrollo de esta investigación.

A mi gran amigo Ricardo Nanfuñay, por todo su apoyo durante este arduo proceso.

Al servicio de oncología del HNAAA, por las facilidades brindadas para el desarrollo de esta investigación.

Al Dr. Wilver Rodríguez por su ayuda brindada.

**El autor**

## RESUMEN

El presente trabajo de investigación tuvo como objetivo determinar un modelo de series de tiempo que pronostique el número de pacientes con cáncer en el Hospital Nacional Almanzor Aguinaga Asenjo, Periodo 2006 – 2017.

El diseño de investigación empleado para este estudio fue un diseño no experimental de tipo longitudinal descriptivo, la población estuvo constituida por todos los registros de las historias clínicas de los pacientes diagnosticados con cáncer en el Hospital Nacional Almanzor Aguinaga Asenjo, así mismo la muestra está conformada por los registros de las historias clínicas de los pacientes diagnosticados con cáncer en el Hospital Nacional Almanzor Aguinaga Asenjo desde enero 2006 hasta diciembre 2017.

Para lograr el objetivo planteado se aplicaron 2 metodologías, la primera fue la metodología de Box y Jenkins para series de tiempo, y la segunda, fue redes neuronales artificiales; con los cuales se determinaron qué modelo se ajusta mejor a los datos, por lo tanto se podrán realizar mejores pronósticos a futuro.

Luego de aplicar la metodología de Box y Jenkins y redes neuronales, se obtuvo un modelo  $SARIMA(2,1,0)(1,1,0)_{12}$ , y un modelo  $NNAR(4,1,3)_{12}$  respectivamente.

Finalmente se llegó a la conclusión que el modelo que mejor pronostica el número de pacientes con cáncer en el Hospital Nacional Almanzor Aguinaga Asenjo. Periodo 2006-2017, es el modelo de redes neuronales  $NNAR(4,1,3)_{12}$ .

**Palabras clave:** Cáncer, pronóstico, series de tiempo, red neuronal.

## ABSTRACT

The objective of this research work was to determine a time series model that predicts the number of cancer patients in the Almanzor Aguinaga Asenjo National Hospital, Period 2006 - 2017.

The research design used for this study was a descriptive longitudinal non-experimental design, the population was made up of all the records of the clinical histories of patients diagnosed with cancer at the Almanzor Aguinaga Asenjo National Hospital, and the sample is made up of for the records of the medical records of patients diagnosed with cancer at the Almanzor Aguinaga Asenjo National Hospital from January 2006 to December 2017.

To achieve the proposed objective, 2 methodologies were applied, the first was the Box and Jenkins methodology for time series, and the second was artificial neural networks; with which it was determined which model best fits the data, therefore better forecasts can be made in the future.

After applying Box and Jenkins methodology and the neural network, a model  $SARIMA(2,1,0)(1,1,0)_{12}$  was obtained, and a model  $NNAR(4,1,3)_{12}$  respectively.

Finally, it was concluded that the model that best fits the number of cancer patients at the Almanzor Aguinaga Asenjo National Hospital. Period 2006-2017, is the model of neural networks  $NNAR(4,1,3)_{12}$ .

**Keywords:** cancer, prognosis, time series, neural network.

## INDICE

RESUMEN .....	v
ABSTRACT.....	vi
INTRODUCCIÓN .....	1
CAPÍTULO I: DISEÑO TEÓRICO .....	3
1.1    Antecedentes .....	3
1.2    Series de tiempo. ....	4
1.2.1    Procesos estocásticos. ....	4
Procesos estocásticos estacionarios. ....	4
Procesos estocásticos no estacionarios. ....	4
1.2.2    Definición de series de tiempo.....	4
1.2.3    Ruido blanco .....	4
1.2.4    Caminata aleatoria .....	5
1.2.5    Componentes de series temporales .....	5
1.2.6    Función de autocovarianzas y de autocorrelación .....	5
Función de autocovarianzas (FACV).....	5
Función de autocorrelación simple .....	5
Función de autocorrelación parcial .....	6
1.2.7    Prueba de raíz unitaria .....	6
La prueba Dickey-Fuller aumentada (DFA) .....	6
1.2.8    Modelos univariantes de series temporales.....	6
Proceso autorregresivo AR(p) .....	6
Proceso de medias móviles MA(q) .....	7
Modelos ARMA(p, q).....	7
Modelo ARIMA(p, d, q) .....	7
Modelo SARIMA(p, d, q)(P, D, Q) .....	8
1.2.9    Metodología para el modelamiento de series temporales .....	9



Identificación del modelo. ....	11
Estimación de los parámetros del modelo. ....	11
Diagnóstico del modelo. ....	12
Prueba de Ljung-Box .....	12
Prueba de Kolmogorov- Smirnov .....	13
Pronóstico. ....	13
1.3    Redes neuronales.....	13
1.3.1    Modelo neuronal artificial.....	14
1.3.2    Neurodinámica.....	15
Entradas.....	15
Salidas .....	15
Regla de propagación.....	15
Función de activación o transferencia.....	16
1.3.3    Tipos de redes neuronales .....	17
El perceptrón.....	17
El perceptrón multicapa .....	17
Backpropagation. ....	18
Variaciones del algoritmo backpropagation. ....	19
1.3.4    Métodos de aprendizaje .....	19
Redes con aprendizaje supervisado .....	20
Redes con aprendizaje no supervisado .....	20
1.3.5    Redes neuronales aplicadas a las series de tiempo .....	20
Modelo NAR caso univariado. ....	20
CAPÍTULO II. MÉTODOS Y MATERIALES.....	21
2.1    Tipo y diseño de investigación.....	21
2.2    Población y muestra .....	21
2.3    Técnica e instrumento de recolección de datos .....	21

2.4	Análisis y procesamiento estadístico de los datos.....	21
CAPÍTULO III. RESULTADOS Y DISCUSIÓN.....		23
3.1	Descripción del comportamiento de la serie número de pacientes con cáncer .....	23
3.2	Estimación del modelo de serie de tiempo para el número de pacientes con cáncer, mediante la metodología de Box y Jenkins.....	33
3.3	Identificación del modelo de serie de tiempo para el número de pacientes con cáncer, mediante la metodología de redes neuronales. ....	82
3.4	Pronóstico del número de pacientes con cáncer para el año 2018, mediante la metodología de Box y Jenkins y redes neuronales .....	85
3.5	Comparación del mejor pronóstico que se ajusta al número de pacientes con cáncer.....	86
CAPÍTULO IV. CONCLUSIONES .....		88
REFERENCIAS BIBLIOGRÁFICAS.....		89
APÉNDICE Y ANEXOS.....		93
ANEXO 1.....		93
ANEXO 2.....		95
ANEXO 3.....		96
ANEXO 4.....		97
ANEXO 5.....		98
ANEXO 6.....		99
ANEXO 7.....		100
ANEXO 8.....		101
ANEXO 9.....		102
ANEXO 10.....		103
ANEXO 11.....		104
ANEXO 12.....		105
ANEXO 13.....		106
ANEXO 14.....		107

ANEXO 15.....	108
ANEXO 16.....	109
ANEXO 17.....	110
ANEXO 18.....	111
ANEXO 19.....	112
ANEXO 20.....	113
ANEXO 21.....	114
ANEXO 22.....	115
ANEXO 23.....	116
ANEXO 24.....	117
ANEXO 25.....	118
ANEXO 26.....	119
ANEXO 27.....	120
ANEXO 28.....	121
ANEXO 29.....	122
ANEXO 30.....	123
ANEXO 31.....	124
ANEXO 32.....	125
ANEXO 33.....	126
ANEXO 34.....	127
ANEXO 35.....	128
ANEXO 36.....	129
ANEXO 37.....	130
ANEXO 38.....	131
ANEXO 39.....	132
ANEXO 40.....	133

## Lista de tablas

Tabla 1. Estadísticos descriptivos del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.....	24
Tabla 2. Test de Dickey-Fuller para la serie del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017 .....	28
Tabla 3. Test de Dickey-Fuller de la primera diferencia al número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.....	32
Tabla 4. Modelos propuestos que explican el comportamiento de la serie de tiempo número de pacientes con cáncer en el HNAAA .....	33
Tabla 5. Test de Ljung-Box de los residuos del modelo $ARIMA(1,1,0)$ .....	36
Tabla 6. Test ANOVA de los residuos del modelo $ARIMA(1,1,0)$ .....	36
Tabla 7. Test de Levene de los residuos del modelo $ARIMA(1,1,0)$ .....	37
Tabla 8. Test de Kolmogorov-Smirnov de los residuos del modelo $ARIMA(1,1,0)$ .....	38
Tabla 9. Test de Ljung-Box de los residuos del modelo $ARIMA(2,1,0)$ .....	41
Tabla 10. Test ANOVA de los residuos del modelo $ARIMA(2,1,0)$ .....	41
Tabla 11. Test de Levene de los residuos del modelo $ARIMA(2,1,0)$ .....	42
Tabla 12. Test de Kolmogorov-Smirnov de los residuos del modelo $ARIMA(2,1,0)$ .....	43
Tabla 13. Test de Ljung-Box de los residuos del modelo $ARIMA(3,1,0)$ .....	46
Tabla 14. Test ANOVA de los residuos del modelo $ARIMA(3,1,0)$ .....	46
Tabla 15. Test de Levene de los residuos del modelo $ARIMA(3,1,0)$ .....	47
Tabla 16. Test de Kolmogorov-Smirnov de los residuos del modelo $ARIMA(3,1,0)$ .....	48
Tabla 17. Test de Ljung-Box de los residuos del modelo $SARIMA(1,1,0)(1,1,0)_{12}$ .....	51
Tabla 18. Test ANOVA de los residuos del modelo $SARIMA(1,1,0)(1,1,0)_{12}$ .....	51
Tabla 19. Test de Levene de los residuos del modelo $SARIMA(1,1,0)(1,1,0)_{12}$ .....	52
Tabla 20. Test de Kolmogorov-Smirnov de los residuos del modelo $SARIMA(1,1,0)(1,1,0)_{12}$ .....	53
Tabla 21. Test de Ljung-Box de los residuos del modelo $SARIMA(1,1,0)(2,1,0)_{12}$ .....	56
Tabla 22. Test ANOVA de los residuos del modelo $SARIMA(1,1,0)(2,1,0)_{12}$ .....	56

Tabla 23. Test de Levene de los residuos del modelo <i>SARIMA</i> (1,1,0)(2,1,0) <sub>12</sub> .....	57
Tabla 24. Test de Kolmogorov-Smirnov de los residuos del modelo <i>SARIMA</i> (1,1,0)(2,1,0) <sub>12</sub> .....	58
Tabla 25. Test de Ljung-Box de los residuos del modelo <i>SARIMA</i> (2,1,0)(1,1,0) <sub>12</sub> .....	61
Tabla 26. Test ANOVA de los residuos del modelo <i>SARIMA</i> (2,1,0)(1,1,0) <sub>12</sub> .....	61
Tabla 27. Test de Levene de los residuos del modelo <i>SARIMA</i> (2,1,0)(1,1,0) <sub>12</sub> .....	62
Tabla 28. Test de Kolmogorov-Smirnov de los residuos del modelo <i>SARIMA</i> (2,1,0)(1,1,0) <sub>12</sub> .....	63
Tabla 29. Test de Ljung-Box de los residuos del modelo <i>SARIMA</i> (2,1,0)(2,1,0) <sub>12</sub> .....	66
Tabla 30. Test ANOVA de los residuos del modelo <i>SARIMA</i> (2,1,0)(2,1,0) <sub>12</sub> .....	66
Tabla 31. Test de Levene de los residuos del modelo <i>SARIMA</i> (2,1,0)(2,1,0) <sub>12</sub> .....	67
Tabla 32. Test de Kolmogorov-Smirnov de los residuos del modelo <i>SARIMA</i> (2,1,0)(2,1,0) <sub>12</sub> .....	68
Tabla 33. Test de Ljung-Box de los residuos del modelo <i>SARIMA</i> (3,1,0)(1,1,0) <sub>12</sub> .....	71
Tabla 34. Test ANOVA de los residuos del modelo <i>SARIMA</i> (3,1,0)(1,1,0) <sub>12</sub> .....	71
Tabla 35. Test de Levene de los residuos del modelo <i>SARIMA</i> (3,1,0)(1,1,0) <sub>12</sub> .....	72
Tabla 36. Test de Kolmogorov-Smirnov de los residuos del modelo <i>SARIMA</i> (3,1,0)(1,1,0) <sub>12</sub> .....	73
Tabla 37. Test de Ljung-Box de los residuos del modelo <i>SARIMA</i> (3,1,0)(2,1,0) <sub>12</sub> .....	76
Tabla 38. Test ANOVA de los residuos del modelo <i>SARIMA</i> (3,1,0)(2,1,0) <sub>12</sub> .....	76
Tabla 39. Test de Levene de los residuos del modelo <i>SARIMA</i> (3,1,0)(2,1,0) <sub>12</sub> .....	77
Tabla 40. Test de Kolmogorov-Smirnov de los residuos del modelo <i>SARIMA</i> (3,1,0)(2,1,0) <sub>12</sub> .....	78
Tabla 41. Pronósticos del número de pacientes con cáncer en el HNAAA con la metodología Box y Jenkins para el año 2018 .....	79
Tabla 42. Comparación de la evaluación de los errores de pronósticos de los modelos propuestos .....	80
Tabla 43. Pesos sinápticos de la primera capa oculta de la red neuronal .....	82

Tabla 44. Pesos sinápticos de la segunda capa oculta de la red neuronal.....	82
Tabla 45. Pesos sinápticos de la tercera capa oculta de la red neuronal.....	83
Tabla 46. Pronósticos del número de pacientes con cáncer en el HNAAA para el año 2018, usando la metodología de Box y Jenkins y redes neuronales .....	85
Tabla 47. Comparación de los modelos de redes neuronales y SARIMA.....	86
Tabla 48. Correlograma del número de pacientes con cáncer en el HNAAA. Chiclayo 2006 - 2017.....	93

## Lista de figuras

Figura 1. Esquema del modelado de series de tiempo .....	10
Figura 2. Modelo Neuronal.....	14
Figura 3. Transformación afín producida por la presencia de la ganancia .....	14
Figura 4. Funciones de activación comunes .....	16
Figura 5. Funcionamiento de un Perceptrón .....	17
Figura 6. Backpropagation Network.....	18
Figura 7. Variaciones del algoritmo backpropagation.....	19
Figura 8. Arquitectura general de un modelo NAR de Redes neuronales .....	20
Figura 9. Número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.....	23
Figura 10. Diagrama de cajas del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.....	25
Figura 11. Función de autocorrelación simple del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.....	26
Figura 12. Función de autocorrelación parcial del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.....	27
Figura 13. Primera diferencia aplicada al número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017 .....	29
Figura 14. Grafica de cajas de la primera diferencia al número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.....	30
Figura 15. Función de autocorrelación simple de la primera diferencia del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.....	31
Figura 16. Función de autocorrelación parcial de la primera diferencia del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.....	31
Figura 17. Residuos del modelo <i>ARIMA</i> (1,1,0) .....	34
Figura 18. Función de autocorrelación de los residuos del modelo <i>ARIMA</i> (1,1,0).....	35
Figura 19. Función de autocorrelación parcial de los residuos del modelo <i>ARIMA</i> (1,1,0)...	35
Figura 20. Histograma de los residuos del modelo <i>ARIMA</i> (1,1,0).....	37

Figura 21. Diagrama Q-Q normal de los residuos del modelo $ARIMA(1,1,0)$ .....	38
Figura 22. Residuos del Pronóstico del modelo $ARIMA(2,1,0)$ .....	39
Figura 23. Función de autocorrelación de los residuos del modelo $ARIMA(2,1,0)$ .....	40
Figura 24. Función de autocorrelación parcial de los residuos del modelo $ARIMA(2,1,0)$ ...	40
Figura 25. Histograma de los residuos del modelo $ARIMA(2,1,0)$ .....	42
Figura 26. Diagrama Q-Q normal de los residuos del modelo $ARIMA(2,1,0)$ .....	43
Figura 27. Residuos del modelo $ARIMA(3,1,0)$ .....	44
Figura 28. Función de autocorrelación de los residuos del modelo $ARIMA(3,1,0)$ .....	45
Figura 29. Función de autocorrelación parcial de los residuos del modelo $ARIMA(3,1,0)$ ...	45
Figura 30. Histograma de los residuos del Pronóstico del modelo $ARIMA(3,1,0)$ .....	47
Figura 31. Diagrama Q-Q normal de los residuos del modelo $ARIMA(3,1,0)$ .....	48
Figura 32. Residuos del modelo $SARIMA(1,1,0)(1,1,0)_{12}$ .....	49
Figura 33. Función de autocorrelación de los residuos del modelo $SARIMA(1,1,0)(1,1,0)_{12}$ .....	50
Figura 34. Función de autocorrelación parcial de los residuos del modelo $SARIMA(1,1,0)(1,1,0)_{12}$ .....	50
Figura 35. Histograma de los residuos del modelo $SARIMA(1,1,0)(2,1,0)_{12}$ .....	52
Figura 36. Diagrama Q-Q normal de los residuos del modelo $SARIMA(1,1,0)(2,1,0)_{12}$ .....	53
Figura 37. Residuos del modelo $SARIMA(1,1,0)(2,1,0)_{12}$ .....	54
Figura 38. Función de autocorrelación de los residuos del modelo $SARIMA(1,1,0)(2,1,0)_{12}$ .....	55
Figura 39. Función de autocorrelación parcial de los residuos del modelo $SARIMA(1,1,0)(2,1,0)_{12}$ .....	55
Figura 40. Histograma de los residuos del modelo $SARIMA(1,1,0)(2,1,0)_{12}$ .....	57
Figura 41. Diagrama Q-Q normal de los residuos del modelo $SARIMA(1,1,0)(2,1,0)_{12}$ .....	58
Figura 42. Residuos del modelo $SARIMA(2,1,0)(1,1,0)_{12}$ .....	59
Figura 43. Función de autocorrelación de los residuos del modelo $SARIMA(2,1,0)(1,1,0)_{12}$ .....	60



Figura 44. Función de autocorrelación parcial de los residuos del modelo $SARIMA (2,1,0)(1,1,0)_{12}$ .....	60
Figura 45. Histograma de los residuos del modelo $SARIMA (2,1,0)(1,1,0)_{12}$ .....	62
Figura 46. Diagrama Q-Q normal de los residuos del modelo $SARIMA (2,1,0)(1,1,0)_{12}$ .....	63
Figura 47. Residuos del modelo $SARIMA (2,1,0)(2,1,0)_{12}$ .....	64
Figura 48. Función de autocorrelación de los residuos del modelo $SARIMA (2,1,0)(2,1,0)_{12}$ .....	65
Figura 49. Función de autocorrelación parcial de los residuos del modelo $SARIMA (2,1,0)(2,1,0)_{12}$ .....	65
Figura 50. Histograma de los residuos del modelo $SARIMA (2,1,0)(2,1,0)_{12}$ .....	67
Figura 51. Diagrama Q-Q normal de los residuos del modelo $SARIMA (2,1,0)(2,1,0)_{12}$ .....	68
Figura 52. Residuos del modelo $SARIMA (3,1,0)(1,1,0)_{12}$ .....	69
Figura 53. Función de autocorrelación de los residuos del modelo $SARIMA (3,1,0)(1,1,0)_{12}$ .....	70
Figura 54. Función de autocorrelación parcial de los residuos del modelo $SARIMA (3,1,0)(1,1,0)_{12}$ .....	70
Figura 55. Histograma de los residuos del modelo $SARIMA (3,1,0)(1,1,0)_{12}$ .....	72
Figura 56. Diagrama Q-Q normal de los residuos del modelo $SARIMA (3,1,0)(1,1,0)_{12}$ .....	73
Figura 57. Residuos del modelo $SARIMA (3,1,0)(2,1,0)_{12}$ .....	74
Figura 58. Función de autocorrelación de los residuos del modelo $SARIMA (3,1,0)(2,1,0)_{12}$ .....	75
Figura 59. Función de autocorrelación parcial de los residuos del modelo $SARIMA (3,1,0)(2,1,0)_{12}$ .....	75
Figura 60. Histograma de los residuos del modelo $SARIMA (3,1,0)(2,1,0)_{12}$ .....	77
Figura 61. Diagrama Q-Q normal de los residuos del modelo $SARIMA (3,1,0)(2,1,0)_{12}$ .....	78
Figura 62. Pronósticos con el modelo $SARIMA (2,1,0)(1,1,0)_{12}$ del número de pacientes con cáncer en el HNAAA para el año 2018 .....	81
Figura 63. Arquitectura de Redes Neuronales Artificiales (RNA) del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.....	83

Figura 64. Pronóstico del número de pacientes con cáncer en el HNAAA, mediante el algoritmo de redes neuronales, para el año 2018 .....	84
Figura 65. Pronóstico del número de pacientes con cáncer en el HNAAA, mediante el modelo SARIMA y RNA, para el año 2018 .....	87

## INTRODUCCIÓN

Uno de los más grandes problemas para el ser humano es tener la capacidad de realizar pronósticos para poder dar una solución a y prevenir futuros problemas sociales, uno de ellos es la salud.

Para poder dar una solución a esta problemática, existen diversos métodos estadísticos, uno de ellos son las series de tiempo, que, a través de la metodología de Box y Jenkins, la cual permite obtener buenos resultados, siempre y cuando sea bien aplicado. Otro de los métodos de pronóstico son las redes neuronales, las cuales también proporcionan excelentes resultados, ya que las redes neuronales pueden identificar relaciones no lineales.

En el sector salud, uno de los principales problemas es el aumento de casos de cáncer, de acuerdo a lo informado por Adolfo Dammert en el 2016 para el diario Gestión, a nivel nacional se presentan cerca de 47 000 casos nuevos de cáncer, mientras que más de 25 000 mueren a causa de esta enfermedad.

En el Hospital Nacional Almanzor Aguinaga Asenjo de Chiclayo, en el servicio de oncología, existe la necesidad de conocer el número aproximado de pacientes con cáncer que ingresarán al servicio de oncología, pero no tiene un modelo que pueda ayudar a estimar dicho número y tomar decisiones que ayuden tanto a la prevención como al tratamiento de dichos pacientes; por lo antes mencionado se planteó el siguiente problema: ¿Cuál es el modelo de series de tiempo que pronostica el número de pacientes con cáncer, Hospital Nacional Almanzor Aguinaga Asenjo; Periodo 2006 – 2017?

En la búsqueda de una solución al problema mencionado anteriormente, el objetivo general de este trabajo es determinar un modelo de series de tiempo para pronosticar el número de pacientes con cáncer, Hospital Nacional Almanzor Aguinaga Asenjo; Periodo 2006 – 2017. Además para una mejor precisión se establecieron los siguientes objetivos específicos: se describió el comportamiento de la serie en estudio, luego se estimó un modelo de series de

tiempo para el número de pacientes con cáncer, mediante la metodología de Box y Jenkins, así mismo se identificó un modelo de series de tiempo para el número de pacientes con cáncer, mediante la metodología de redes neuronales, también se pronosticó el número de pacientes con cáncer para el año 2018, mediante la metodología de Box y Jenkins y redes neuronales, y finalmente se comparó el mejor modelo que se ajusta al número de pacientes con cáncer.

Se tiene como hipótesis que el modelo para pronosticar el número de pacientes con cáncer, Hospital Nacional Almanzor Aguinaga Asenjo; Periodo 2018, será un modelo de redes neuronales ya que proporcionará mejores pronósticos, además que tendrá menor error que el modelo obtenido mediante la metodología de Box y Jenkins.

La presente investigación se justifica porque no existe en el medio investigaciones que pronostiquen el número de pacientes con cáncer; por lo tanto este modelo es un valioso aporte que ayudará a tomar decisiones óptimas en el ámbito del conocimiento y la prevención por parte de las autoridades de EsSalud dado a que existen pocos estudios realizados relacionados con el presente tema, y los que existen son aplicados sobre todo a la descripción de los diferentes tipos de cáncer.

# CAPÍTULO I: DISEÑO TEÓRICO

## 1.1 Antecedentes

**Siegel, Miller y Jemal (2017)**, realizaron la investigación “Cancer Statistics, 2017”, para la revista científica CA: A Cancer Journal for Clinicians; en la que concluyeron que: Con un método de proyección temporal (el modelo vectorial autorregresivo) se pronosticó para el año 2017 un total de 1’688,780 casos nuevos en Estados Unidos.

**Ortiz (2012)**, en su investigación “Modelos Econométricos y de Redes Neuronales para predecir la Oferta Maderera en México: ARIMA vs NAR y ARMAX vs NARX”, concluye que las Redes Neuronales Artificiales (RNA) pueden ser una muy buena herramienta para los econométristas, ya que pronostican bien, rápidamente, y son muy fáciles de usar, dando para el caso de los pronósticos tan buenos resultados como los obtenidos mediante métodos econométricos tradicionales.

**Gómez (2010)**, en su estudio “Pronóstico de una serie temporal usando redes neuronales”; afirma que las redes neuronales proporcionan buenas predicciones comparadas con las predicciones del modelo de serie de tiempo.

**Cruz (2007)**, realizó la investigación “Pronósticos en el mercado de derivados utilizando redes neuronales y modelos ARIMA”, concluye que la principal aportación es demostrar que el modelo propuesto basado en redes neuronales es capaz de obtener buenas aproximaciones tanto en el ajuste como en el pronóstico, además de observar que, en este caso en particular, la aproximación resulta mejor que la generada por medio de la metodología ARIMA.

## **1.2 Series de tiempo.**

### **1.2.1 Procesos estocásticos.**

Desde la perspectiva de Uriel y Peiró (2000) definen a los procesos estocásticos “Como una familia de variables aleatorias que corresponden a momentos sucesivos del tiempo. Sera designado por  $Y(t, u)$ , donde  $t$  es el tiempo y  $u$  es la variable aleatoria” (p. 18).

Gujarati y Porter (2010) consideran a los procesos estocásticos como estacionarios o no estacionarios:

#### **a) Procesos estocásticos estacionarios.**

“Son aquellos cuya media y varianza son constantes en el tiempo y si la covarianza entre dos periodos depende sólo de la distancia entre estos dos periodos, y no del tiempo en el cual se calculó la covarianza”. (p. 740)

#### **b) Procesos estocásticos no estacionarios.**

“Como aquellos cuya distribución de probabilidad varia de forma no constante. Hay dos tipos de caminatas aleatorias: Caminata aleatoria sin deriva o sin desvío (sin término constante), y caminata aleatoria con deriva o con desvío (hay un término constante)”. (p. 741)

### **1.2.2 Definición de series de tiempo**

Kozhan (2010) considera que una serie de tiempo “es una secuencia de datos numéricos en la cual las observaciones se miden en un instante de tiempo particular (anual, trimestral, mensual, diaria, etc.); además, su objetivo principal es estudiar la dinámica de los datos”. (p. 54)

### **1.2.3 Ruido blanco**

Según Novales (2010), al “Ruido blanco se le conoce a la sucesión de variables aleatorias con promedio igual a cero, igual varianza, e independientes en el tiempo.” (p. 414)

#### 1.2.4 Caminata aleatoria

La caminata aleatoria de acuerdo a Novales (2010), “es un proceso estocástico  $\{Y_t\}$  cuyas primeras diferencias forman un proceso de ruido blanco”, es decir:

$$\nabla Y_t = \varepsilon_t$$

O

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

(pp. 414-415)

#### 1.2.5 Componentes de series temporales

Son movimientos o variaciones características que pueden medirse y observarse por separado. Gómez (2010) afirma que “los componentes son tendencia, es el crecimiento o declinación de la serie; ciclo, oscilaciones alrededor de la tendencia; variación estacional, son patrones periódicos que se repiten año tras año; fluctuaciones irregulares, son movimientos erráticos que siguen un patrón indefinido”. (p. 5)

#### 1.2.6 Función de autocovarianzas y de autocorrelación

##### a) Función de autocovarianzas (FACV)

Novales (2000) determina que la función de autocovarianzas “es una función, a la que en lo sucesivo nos referimos como FAC, que para instante  $t$  y cada número entero  $k$  toma un valor, denotado por  $\gamma_k(t)$ , igual a la covarianza entre  $y_t$  e  $y_{t-k}$ , es decir:  $\gamma_k(t) = Cov(Y_t, Y_{t-k})$ ”. (p. 416)

##### b) Función de autocorrelación simple

Novales (2000), tiene en cuenta que “la Función de Autocorrelación Simple (FAS), es aquella que para cada instante  $t$  y cada entero  $k$  toma un valor  $\rho_k(t)$  igual a la correlación entre

$$y_t \text{ e } y_{t-k}: \rho_k(t) = \frac{Cov(Y_t, Y_{t-k})}{\sqrt{Var Y_t} \sqrt{Var Y_{t-k}}} = \frac{\gamma_k(t)}{\sqrt{Var Y_t} \sqrt{Var Y_{t-k}}}$$
 (p. 416)

### c) Función de autocorrelación parcial

Novales (2000), especifica que la Función de Autocorrelación Parcial (FAP) “es aquella para cada instante  $t$  y cada entero  $k$  toma un valor igual a la correlación entre  $Y_t$  e  $Y_{t-k}$  ajustada por el efecto de los retardos intermedios  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k-1}$ ”. (p. 417)

#### 1.2.7 Prueba de raíz unitaria

Carter, Griffiths y Lim (2011) consideran que “hay muchas pruebas de raíz unitaria, las cuales sirven para determinar si una serie es estacionaria o no. La más popular es la de **Dickey-Fuller**”. (p. 484)

Gujarati y Porter (2010) estipulan que **la prueba Dickey-Fuller aumentada (DFA)** “es unidireccional porque la hipótesis alternativa es:

$$H_1: \delta < 0 \text{ (Es decir, la serie de tiempo es estacionaria)}.”.$$

Además afirman que consiste en estimar la siguiente regresión:

$$\Delta Y_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \sum_{i=1}^m \alpha_i \Delta Y_{t-i} + \varepsilon_t,$$

donde  $\varepsilon_t$  es un término de error puro de ruido blanco, y donde:

$$\Delta Y_{t-1} = (Y_{t-1} - Y_{t-2}), \Delta Y_{t-2} = (Y_{t-2} - Y_{t-3}), \text{ etc}.”. (p. 756-757)$$

#### 1.2.8 Modelos univariantes de series temporales

##### a) Proceso autorregresivo AR(p)

Según Cowpertwait y Metcalfe (2009), definen que la serie  $\{Y_t\}$  es un proceso autorregresivo de orden  $p$ , o simplemente  $AR(p)$ , si

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t$$



Donde  $\{\varepsilon_t\}$  es ruido blanco y  $\alpha_i$  son los parámetros del modelo con  $\alpha_p \neq 0$  para un proceso de orden  $p$ . Se puede expresar como un polinomio de orden  $p$  en términos del operador de desplazamiento hacia atrás:

$$\theta_p(B)x_t = (1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_p B^p)Y_t = \varepsilon_t$$

(p. 79)

### **b) Proceso de medias móviles $MA(q)$**

Está definido, según Shumway, Stoffer (2011) como:

$$x_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q}$$

Donde hay  $q$  rezagos en la media móvil y  $\beta_1, \beta_2, \dots, \beta_q$  ( $\beta_q \neq 0$ ) son parámetros. Aunque no es necesario, suponemos que  $w_t$  es una serie de ruido blanco con media cero y varianza  $\sigma_\varepsilon^2$ , a menos que se indique lo contrario. (p. 90)

### **c) Modelos $ARMA(p, q)$**

Son modelos mixtos y se representa por medio de la ecuación:

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q}$$

Entonces,  $x_t$  sigue un modelo de promedio móvil autorregresivo, o  $ARMA(p, q)$ , donde  $p$  y  $q$  son las órdenes de las partes promedio autoregresivas y móviles, respectivamente. (Prado, West, 2010, p. 65)

### **d) Modelo $ARIMA(p, d, q)$**

Guzmán (2014), afirma que la condición de serie estacionaria requiere estabilidad en la media y la varianza; sin embargo en la vida real los procesos no satisfacen estos requerimientos. Un proceso no estacionario en varianza es tratado por las transformaciones de Box-Cox, las cuales son definidas como  $v_t = (z_t^\lambda - 1)/\lambda$  para  $\lambda \neq 0 \in \mathbb{R}$ . Para un modelo dado, el valor óptimo de  $\lambda$  se encuentra mediante la minimización de la suma del error cuadrático del

modelo. En caso de que  $\lambda$  se encuentre cercano o igual cero, se utiliza una transformación con el logaritmo natural  $v_t = \ln(z_t)$ . (p. 16)

Si el proceso no estacionario es el resultado de una media variable, el proceso puede ser diferenciado mediante  $v_t = (1 - L)^d z_t$ , ajustando el orden de diferenciación  $d = 1$  o  $d = 2$  empleados comúnmente para inducir a una media constante, el modelo  $ARMA(p, q)$  para el proceso diferenciado  $v$  es referido como un modelo  $ARIMA(p, d, q)$  para el proceso  $z$ . (p. 16)

Un modelo  $ARIMA(p, d, q)$  responde a la siguiente ecuación:

$$\phi_p(L)\nabla^d Y_t = c + \theta_q(L)\varepsilon_t$$

Donde  $\nabla^d$  es el operador de diferencias de orden  $d$ , definido en general como:

$$\nabla^d z_t = (1 - L)^d z_t = z_t - z_{t-d}$$

(p. 16)

#### e) Modelo $SARIMA(p, d, q)(P, D, Q)$

Chatfield y Xing (2019) manifiestan que en la práctica, muchas series de tiempo contienen un componente periódico estacional, que repite las observaciones de cada  $s$ . Por ejemplo, con observaciones mensuales, donde  $s = 12$ , normalmente podemos esperar que  $X_t$  dependa de valores en rezagos anuales, como  $X_{t-12}$ , y quizás  $X_{t-24}$ , así como de valores no estacionales más recientes como  $X_{t-1}$  y  $X_{t-2}$ . Box y Jenkins (1970) generalizaron el modelo  $ARIMA$  para tratar la estacionalidad y definieron un modelo  $ARIMA$  general multiplicativo estacional ( $SARIMA$ ) como

$$\phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)Z_t,$$

donde  $B$  denota el operador de desplazamiento hacia atrás,  $\phi_p, \Phi_P, \theta_q, \Theta_Q$  son polinomios de orden  $p, P, q, Q$ , respectivamente,  $Z_t$  denota un proceso puramente aleatorio y

$$W_t = \nabla^d \nabla_s^D X_t$$

denota la serie diferenciada. Si el entero  $D$  no es cero, entonces está involucrada la diferenciación estacional. El modelo anterior se denomina modelo SARIMA de orden  $(p, d, q) \times (P, D, Q)_s$ . (p. 103)

El modelo SARIMA parece bastante complicado a primera vista, así que veamos los diferentes componentes del modelo uno por uno. Primero, la serie diferenciada  $\{W_t\}$  se forma a partir de la serie original  $\{X_t\}$  mediante la diferenciación apropiada para eliminar los términos no estacionarios. Si  $d$  es distinto de cero, entonces hay una diferenciación simple para eliminar la tendencia, mientras que las diferencias estacionales  $\nabla_s$  pueden usarse para eliminar la estacionalidad. Por ejemplo, si  $d = D = 1$  y  $s = 12$ , entonces

$$W_t = \nabla \nabla_{12} X_t = \nabla_{12} X_t - \nabla_{12} X_{t-1} = (X_t - X_{t-12}) - (X_{t-1} - X_{t-13})$$

En la práctica, los valores de  $d$  y  $D$  suelen ser cero o uno, y rara vez dos. (p. 103)

Ahora veamos el término  $AR$  estacional,  $\Phi_P(B^s)$ . Supongamos, para simplificar, que  $P = 1$ . Entonces  $\Phi_1(B^s)$  tendrá la forma  $(1 - C - B^s)$ , donde  $C$  denota una constante, lo que simplemente significa que  $W_t$  dependerá de  $W_{t-s}$ , ya que  $(B^s)W_t = W_{t-s}$ . De manera similar, un término de  $MA$  estacional de orden uno significa que  $W_t$  dependerá tanto de  $Z_{t-s}$  como de  $Z_t$ . (p. 103)

### 1.2.9 Metodología para el modelamiento de series temporales

Gujarati y Porter (2010) especifican que “La metodología Box y Jenkins resulta útil para modelar series de tiempo. El método considera cuatro pasos: (p. 777)

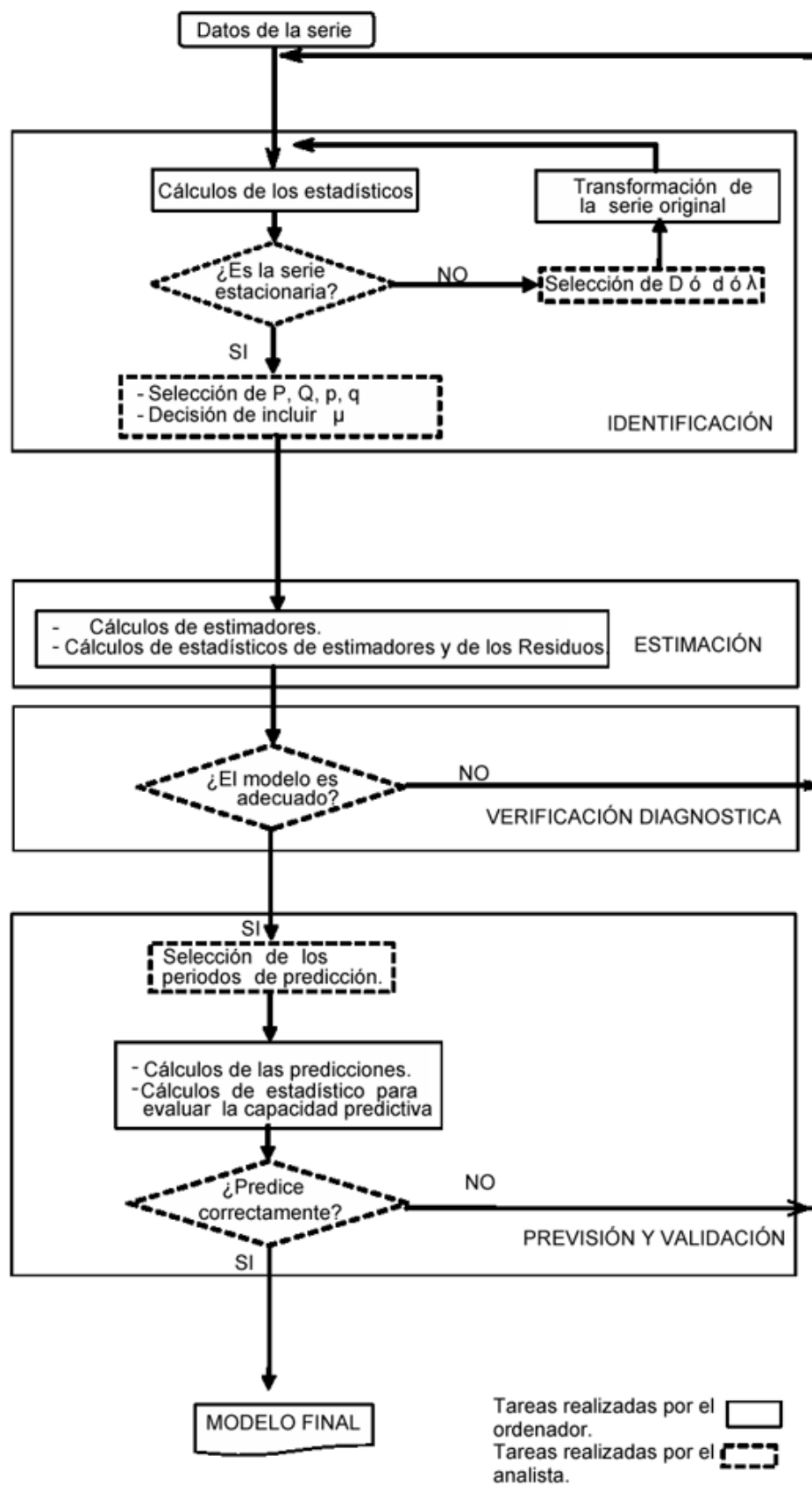


Figura 1. Esquema del modelado de series de tiempo. (Uriel y Peiró (2000))

**a) Identificación del modelo.**

Uriel y Peiró (2000) precisan: “Primero procede a efectuar un análisis de estacionariedad de la serie. En el caso de que se trate de una serie no estacionaria se aplican las transformaciones adecuadas al objeto de convertirla en estacionaria”. Así, si se dispone de una serie  $Y_t$ , se elegirán unos valores de  $d$  y  $\lambda$  de forma que la serie:

$$\varepsilon_t = (1 - L)^d Y_t^{(\lambda)}$$

goce de la propiedad de estacionaridad. Tanto a  $Y_t$  como a  $\varepsilon_t$  se le puede restar el valor de una media si se considera necesario. (pp. 65 - 66)

“En la segunda etapa de la fase se procede a determinar el orden de la parte autorregresiva y el orden de la parte de medias móviles del proceso ARMA que se considere haya podido generar a la serie estacionaria”. (Uriel y Peiró ,2000. pp. 65 - 66)

**b) Estimación de los parámetros del modelo.**

Uriel y Peiró (2000), afirman que en esta fase, como es lógico, el objetivo es la obtención de estimadores de los parámetros  $\alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q, \mu_\varepsilon$  y  $\sigma_\varepsilon^2$ . Para designar a los parámetros autorregresivos y de medias móviles se utilizará el vector

$$\hat{\beta} = (\alpha_1, \alpha_2, \dots, \alpha_p; \beta_1, \beta_2, \dots, \beta_q)' \text{ (p. 108)}$$

Los supuestos que se adoptarán en la exposición del proceso de estimación son los siguientes:

- i.  $\varepsilon \sim N(0, \sigma_\varepsilon^2 I)$ .
- ii. El proceso  $\varepsilon_t$  es estacionario.
- iii. El proceso  $\varepsilon_t$  es invertible.
- iv.  $\delta_\varepsilon$  y, en consecuencia,  $\mu_\varepsilon$  son nulos. (p. 108)

Para realizar la estimación se dispone de una muestra de tamaño  $T$  de la variable  $Y$ . Al tomar diferencias de orden  $d$ , la variable  $\varepsilon_t$  constará de  $T - d$  elementos. Con objeto de

simplificar la notación, a la diferencia  $T - d$  se le designará por  $N$ , y a los elementos de la muestra de  $\varepsilon_t$  se les designará por el vector

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)' \text{ (p. 108)}$$

Si se adopta el criterio mínimo – cuadrático, los estimadores se obtienen minimizando la expresión:

$$S = \sum \varepsilon_t^2 \text{ (p. 108)}$$

Si se adopta el criterio máximo – verosímil, entonces los estimadores se obtendrán mediante maximización de la función de veromilitud

$$L = f(\varepsilon_1, \dots, \varepsilon_N) \text{ (pp. 108)}$$

Cualquiera que sea el criterio adoptado, en la estimación de un modelo ARMA se plantea por una parte el problema de determinación de los valores iniciales y por otra parte el problema de que, en general, los modelos de este tipo son no lineales. (p. 109)

#### **c) Diagnóstico del modelo.**

Uriel y Peiró (2000) puntualizan que “Un modelo ideal sería si: sus residuos se comportan como ruido blanco, si es estacionario, los coeficientes son estadísticamente significativos y suficientes para representar la serie, el grado de ajuste es elevado en comparación al de otros modelos”. (p. 127).

#### **d) Prueba de Ljung-Box**

Villavicencio (2007) afirma que esta prueba “permite probar en forma conjunta de que todos los coeficientes de autocorrelación son independientes, está definida como:

$$LB = n(n+2) \sum_{k=1}^m \left( \frac{\hat{\rho}_k^2}{n-k} \right) \sim \chi_{(m)}^2$$

Donde

$n$  = tamaño de la muestra,  $m$  longitud del rezago.

$k$  = número de rezagos que se analizan ( $k = 12, 24, 36, 48$ ). (p. 5)

**e) Prueba de Kolmogorov- Smirnov**

Siegel y Castellan (1998) definen que “es una prueba estadística no paramétrica, la cual determina si las puntuaciones en una muestra pueden razonablemente provenir de una población que tiene una distribución teórica”. (p. 75)

**f) Pronóstico.**

Hanke y Wichern (2010) precisan que “una vez que se ha encontrado un modelo adecuado, es factible elaborar los pronósticos de uno o varios periodos futuros. Con base en los pronósticos también se pueden construir intervalos de predicción”. (p. 411)

Es buena idea hacer un seguimiento de los errores de pronóstico, usando los siguientes indicadores:

- Error cuadrático medio (RMSE), Error medio absoluto (MAE) y Error absoluto medio porcentual (MAPE).
- Solo para el caso de la metodología de Box y Jenkins se calcularán los coeficientes de información de Akaike (AIC) y el bayesiano (BIC).

**1.3 Redes neuronales.**

Según Khare y Nagendra (2007), definen que la red neuronal es un sistema de red en paralelo a gran escala que consiste en un bloque básico llamado "neurona". Estas neuronas están interconectadas en capas individuales o múltiples. Éstos están biológicamente inspirados, es decir, están compuestos de elementos que actúan de manera análoga a las funciones más elementales de las neuronas biológicas. Estos elementos se organizan entonces de una manera que puede (o no) estar relacionada con la anatomía del cerebro. A pesar de este parecido superficial, las RNA exhiben una serie de características cerebrales. (p. 26)

### 1.3.1 Modelo neuronal artificial

Haykin (1999), manifiesta que “existen tres elementos básicos del modelo neuronal: un conjunto de sinapsis o enlace de conexión, un sumador para sumar las señales de entrada y una función de activación para limitar la amplitud de la salida de una neurona”. (p. 33)

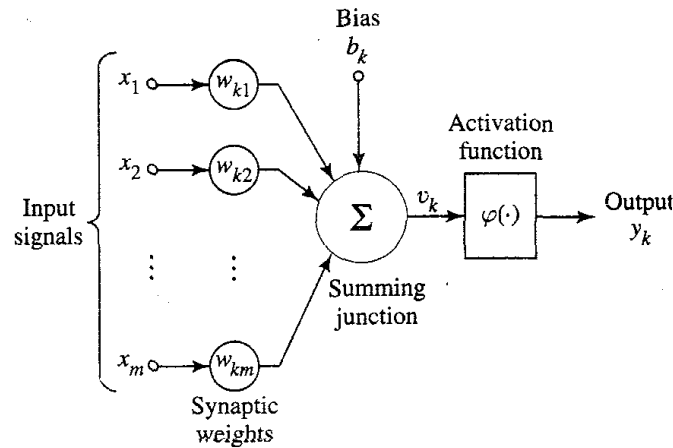


Figura 2. Modelo Neuronal. (Haykin, 1999)

El uso de la ganancia  $b_k$  tiene el efecto de aplicar una transformación afín a la salida de la combinación lineal del modelo mostrado en la Figura 2, el cual es llamado “entrada neta de la red”; la cual, dependiendo de la ganancia, la relación entre el campo inducido local o el potencial de activación de la neurona y de la combinación lineal de la salida es modificada en la forma que se ilustra en la Figura 3. (Haykin, 1999, p. 33)

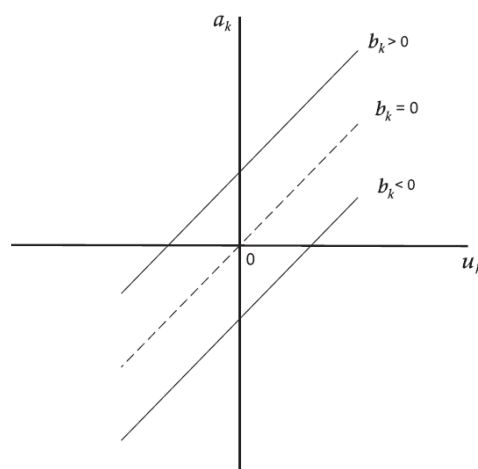


Figura 3. Transformación afín producida por la presencia ganancia ( $b_k$ ). (Gómez, 2010)



### 1.3.2 Neurodinámica

- **Entradas**

La capa de entrada, “funciona como un regulador de las entradas, transfiriendo datos a la siguiente capa. Esto puede requerir un pre procesamiento para que la red sólo maneje datos numéricos”. (Cruz, 2007, p. 20)

- **Salidas**

Cruz (2007) considera que la capa de salida “funciona de una forma similar a la capa de entrada excepto que transfiere la información de la red al exterior”. (p. 20)

- **Regla de propagación**

Desde la perspectiva de Cruz (2007), la regla de propagación “permite calcular el valor del potencial post-sináptico  $h_i$  de la neurona a partir de las entradas y pesos correspondientes. La cual es la suma ponderada de entradas y pesos sinápticos”. (p. 21)

$$h_i(t) = \sum_j w_{ij}x_j$$

Una regla de propagación no lineal, “involucra la multiplicación de todas las entradas de la neurona, lo cual se acerca mucho a la funcionalidad de la neurona biológica, pero aumenta su complejidad tanto en su análisis. Se denomina neurona sigma-pi, por emplear sumas y multiplicaciones”. (Cruz, 2007, p. 21)

$$h_i(t) = \sum_{j_1, j_2, \dots, j_p} w_{ij_1, j_2, \dots, j_p} x_{j_1, j_2, \dots, j_p}$$

Otra regla de propagación empleada es la distancia euclídea, la cual, “determina la distancia entre el vector de entradas y el de pesos. Cuando los vectores de entrada y de pesos son similares, la distancia es muy pequeña, y cuando son diferentes la distancia se incrementa”. (Cruz, 2007, p. 22)

$$h_i^2(t) = \sum_j (x_j - w_{ij})^2$$

- **Función de activación o transferencia**

Cruz (2007) explica que “esta función determina el estado de activación actual  $a_i(t)$  a partir del potencial postsináptico  $h_i(t)$  y del estado de activación anterior  $a_i(t - 1)$ ”. (p. 22)

$$a_i(t) = f_i(a_i(t - 1), h_i(t))$$

“La función identidad, la función escalón y la función lineal, que son las más sencillas. Aunque dependiendo del algoritmo de aprendizaje, se requiere que la función de activación sea derivable, las más empleadas son las funciones del tipo sigmoideo”. (Cruz, 2007, p. 22)

Además, “La función gaussiana es utilizada con reglas de propagación que involucran el cálculo del cuadrado de las distancias entre vectores de entrada y pesos. Las funciones sinusoidales se emplean en casos en que es necesario expresar una periodicidad temporal”. (Cruz, 2007, p. 23)

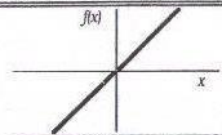
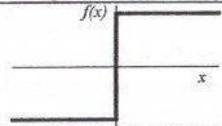
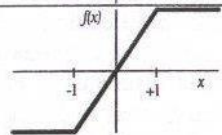
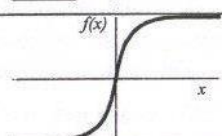
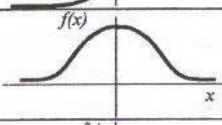
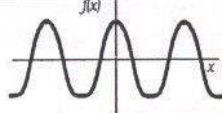
	Función	Rango	Gráfica
<b>Identidad</b>	$y = x$	$[-\infty, +\infty]$	
<b>Escalón</b>	$y = \text{sign}(x)$ $y = H(x)$	$\{-1, +1\}$ $\{0, +1\}$	
<b>Lineal a tramos</b>	$y = \begin{cases} -1, & \text{si } x < -l \\ x, & \text{si } -l \leq x \leq l \\ +1, & \text{si } x > l \end{cases}$	$[-1, +1]$	
<b>Sigmoidea</b>	$y = \frac{1}{1 + e^{-x}}$ $y = \text{tgh}(x)$	$[0, +1]$ $[-1, +1]$	
<b>Gaussiana</b>	$y = Ae^{-Bx^2}$	$[0, +1]$	
<b>Sinusoidal</b>	$y = A \text{sen}(\omega x + \varphi)$	$[-1, +1]$	

Figura 4. Funciones de activación comunes. (L’Huillier, G., & Weber, R., 2010)

### 1.3.3 Tipos de redes neuronales

Ortiz (2012) afirma que “existen diversos tipos de modelos de redes neuronales, tales como: perceptrón, perceptrón multicapa, propagación hacia atrás (backpropagation), entre otros”. (pp. 80-81)

- **El perceptrón**

Kriesel (2007) asiente que Rosenblatt en 1957, definió la suma ponderada ya discutida y una función de activación no lineal como componentes del perceptrón. (p. 71)

No existe una definición establecida para un perceptrón, pero la mayoría de las veces el término percepción se usa para describir una red de retroalimentación con conexiones de acceso directo. (p. 71)

Esta red tiene una capa de neuronas de exploración (retina) con conexiones ponderadas estáticamente a la siguiente capa siguiente y se llama capa de entrada; pero se permite cambiar los pesos de todas las otras capas. Todas las neuronas subordinadas a la retina son detectores de patrones. (p. 71)

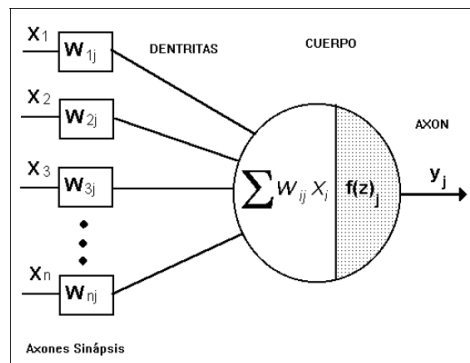


Figura 5. Funcionamiento de un Perceptrón. (Ortiz, 2012)

- **El perceptrón multicapa**

Alsmadi, Omar y Noah (2009), definen que este tipo de perceptrón es una modificación del perceptrón lineal estándar porque utiliza tres o más capas de neuronas (nodos) con funciones de activación no lineales y es más potente que el perceptrón; además de ser

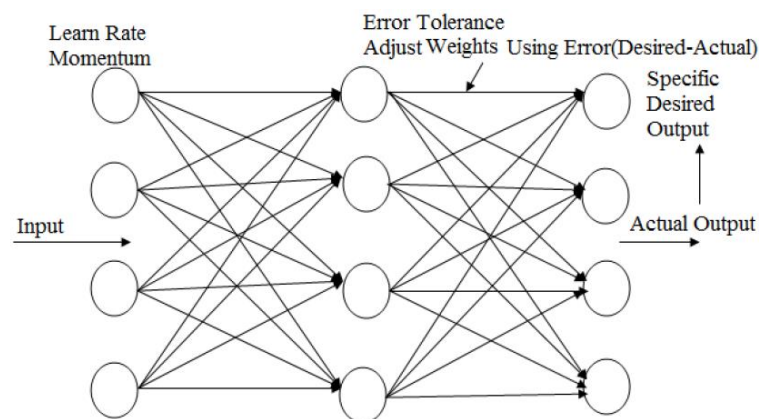
modelos no lineales, flexibles y de propósito general, que constan de varias unidades organizadas en múltiples capas. La complejidad de la red MLP se puede cambiar variando el número de capas y el número de unidades en cada capa. (p. 378)

- **Backpropagation.**

Alam, Kumar, Chowdhury, Siddique (2015), afirman que es una forma de aprendizaje supervisado para redes multicapa, también conocida como la regla delta generalizada. Los datos de error en la capa de salida se propagan nuevamente a los anteriores, lo que permite actualizar los pesos entrantes de estas capas. Se usa con mayor frecuencia como algoritmo de entrenamiento en aplicaciones actuales de redes neuronales. El algoritmo de propagación de retorno es una herramienta matemática involucrada; sin embargo, la ejecución de las ecuaciones de entrenamiento se basa en procesos iterativos, y por lo tanto es fácilmente implementable en una computadora. (p. 271)

Durante la sesión de entrenamiento de la red, un par de patrones se llama patrón de entrada y patrón deseado. El patrón de entrada provoca respuestas de salida en cada neurona en cada capa y, por lo tanto, una salida aceptable en la capa de salida. En la capa de salida, la diferencia entre las salidas reales y las producidas produce una señal de error. Esta señal de error depende de los valores de los pesos de las neuronas en cada capa. (Alam, et al., 2015, p.

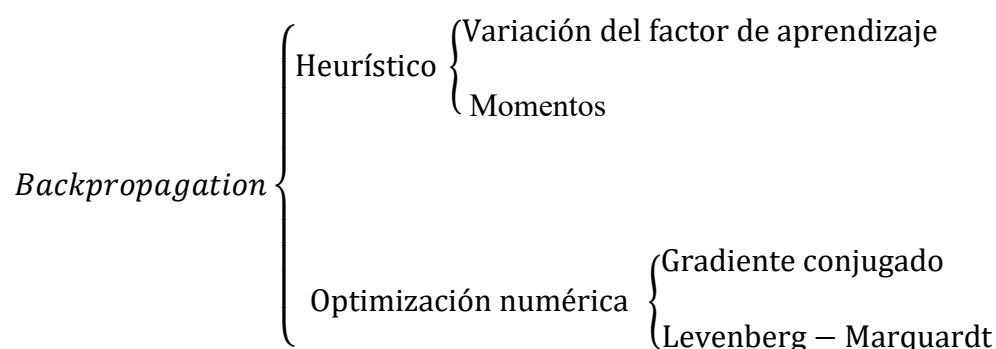
271)



*Figura 6. Backpropagation Network.*

- **Variaciones del algoritmo backpropagation.**

Gómez (2010), considera que “existen dos categorías: las técnicas heurísticas, las cuales incluyen variación del factor de aprendizaje y momentos; y las técnicas de optimización numérica, que son variaciones del método de Newton como el gradiente conjugado y el algoritmo de Levenberg-Marquardt”. (p. 67)



*Figura 7. Variaciones del algoritmo backpropagation*

#### **1.3.4 Métodos de aprendizaje**

Khare y Nagendra (2007), definen que un proceso de "aprendizaje", o "formación", forma las interconexiones (correlaciones) entre las neuronas. Se logra proporcionando datos de entrada y salida conocidos de una manera ordenada a la red neuronal. (p. 35)

El aprendizaje corresponde a los cambios de parámetros (peso). Una red puede ser sometida a un aprendizaje "supervisado" o "sin supervisión". En el caso anterior, los prototipos externos se utilizan como salidas "objetivo" para entradas específicas, y se le da a la red un algoritmo de aprendizaje para seguir y calcular nuevos pesos de conexión que acercan la salida a la salida "objetivo". El aprendizaje sin supervisión es el tipo de aprendizaje que se lleva a cabo sin un "maestro". En el aprendizaje sin supervisión, la respuesta deseada no se conoce, por lo tanto, la información de error explícita no se puede utilizar para mejorar el comportamiento de la red. (p. 35)

- **Redes con aprendizaje supervisado**

Cárdenas (2015) puntualiza: “Proporcionan un conjunto de datos de entrada y la respuesta correcta. Entonces se compara la respuesta calculada de la red y se ajustan los pesos para conseguir una respuesta correcta. Es útil para regresión y clasificación”. (p. 35)

- **Redes con aprendizaje no supervisado**

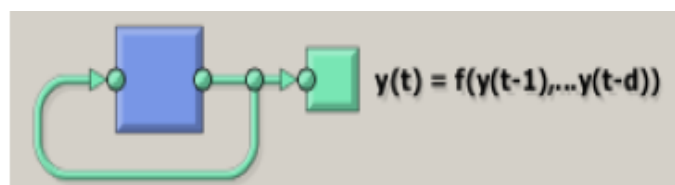
Las redes con aprendizaje no supervisado, de acuerdo a lo referido por Cárdenas (2015), “tratan de imitar el comportamiento humano, la capacidad de aprender sin supervisión explícita. La red debe auto-organizarse y debe actuar ante algunos aspectos del conjunto de entrada como la existencia de redundancia o grupos en el conjunto de datos”. (p. 35)

### **1.3.5 Redes neuronales aplicadas a las series de tiempo**

Según Mercado, Pedraza y Martínez (2015), los modelos de redes neuronales son más robustas y eficaces para la predicción y el modelado de series de tiempo; para ello debe configurarse de tal manera que las entradas de la red sean valores pasados al que se desea predecir; para entrenar la red se deben ajustar los pesos para obtener los valores deseados de la serie en un instante de tiempo específico. (p. 92.)

- **Modelo NAR caso univariado.**

Ortiz (2012) manifiesta que el Modelo NAR es “es un modelo neuronal autorregresivo no lineal, basado en una única serie de datos; que predice valores de la serie utilizando las observaciones pasadas. Es considerada una de las mejores para analizar y predecir información de series de tiempo”. (p. 87). La arquitectura general se presenta a continuación:

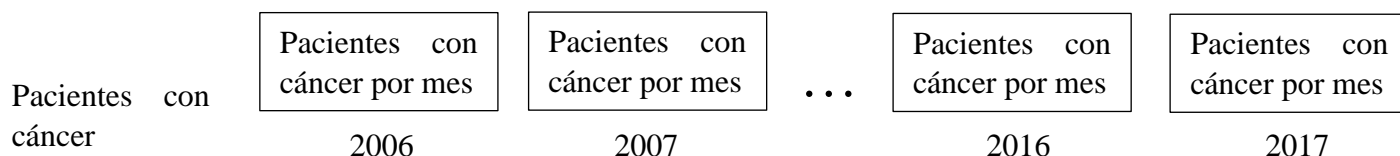


*Figura 8.* Arquitectura general de un modelo NAR de Redes neuronales. (Ortiz, 2012)

## CAPÍTULO II. MÉTODOS Y MATERIALES

### 2.1 Tipo y diseño de investigación

El diseño de investigación empleado, fue un diseño no experimental de tipo longitudinal descriptivo.



### 2.2 Población y muestra

#### Población

La población está constituida por todos los registros de las historias clínicas de los pacientes diagnosticados con cáncer en el Hospital Nacional Almanzor Aguinaga Asenjo.

#### Muestra

La muestra está conformada por los registros de las historias clínicas de los pacientes diagnosticados con cáncer en el Hospital Nacional Almanzor Aguinaga Asenjo desde enero 2006 hasta diciembre 2017.

### 2.3 Técnica e instrumento de recolección de datos

La técnica de recolección de datos fue la observación documental de las historias clínicas de los pacientes. Y el instrumento de recolección de datos fue el formato automatizado “Registro de aspectos clínicos de los pacientes diagnosticados con cáncer en el Hospital Nacional Almanzor Aguinaga Asenjo” elaborado por el investigador (Ver anexo 01).

### 2.4 Análisis y procesamiento estadístico de los datos

El análisis estadístico consistió en primer lugar en seguir los pasos de la metodología de Box y Jenkins, la cual consistió en identificar el modelo, estimar los parámetros del modelo elegido, examinar el diagnóstico obtenido y finalmente se realiza el pronóstico.

Para el proceso de identificación se graficó la serie de tiempo, el correlograma total; y parcial, así mismo se realizaron la prueba de Dickey – Fuller para verificar la estacionariedad de la serie.

Para el diagnostico se utilizó las pruebas de Kolmogorov Smirnov con un nivel de confianza del 95%, para determinar si los residuales siguen una distribución normal y es una variable de ruido blanco. Para el proceso de estimación, se estimaron los posibles modelos que se ajustan a la serie analizada y finalmente se calcularon los pronósticos de la serie número de pacientes con cáncer en el Hospital Nacional Almanzor Aguinaga Asenjo para el año 2018.

Por otro lado se utilizó la metodología de redes neuronales, la cual consistió en encontrar los pesos óptimos para determinar las mejores estimaciones.

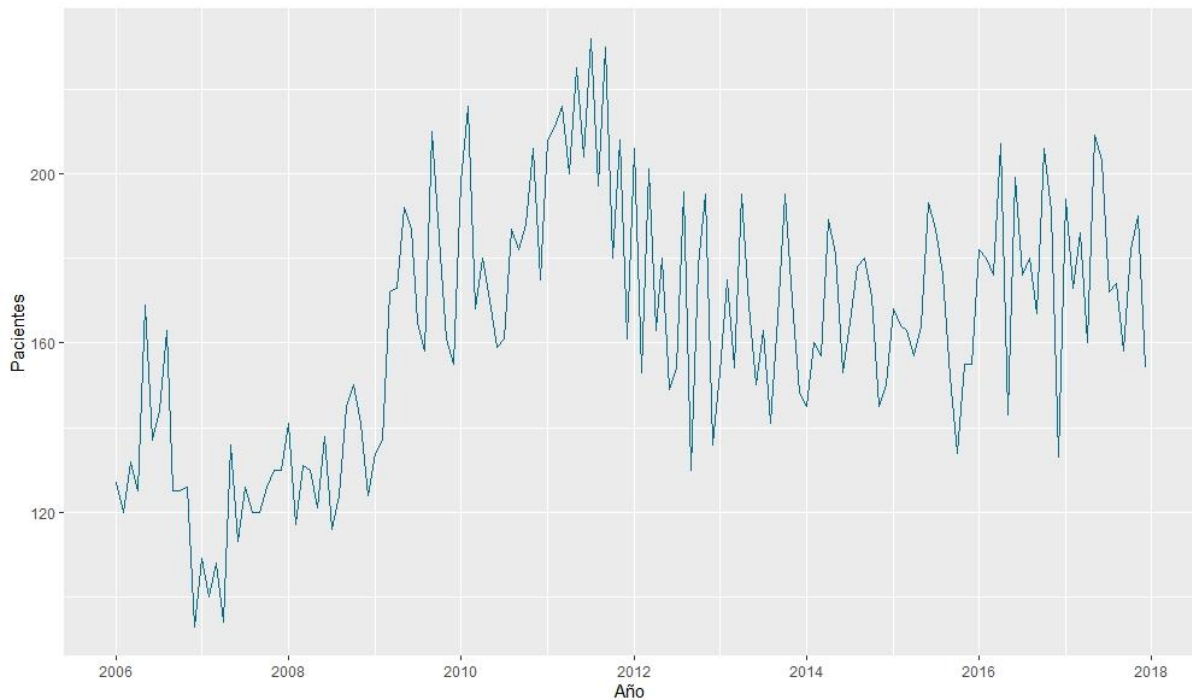
Posteriormente se realizaron los análisis comparativos de los pronósticos mediante los métodos de Box y Jenkins y redes neuronales; de la serie de datos utilizados, para ello se calculó el RMSE y el MAPE.

Para el procesamiento y análisis de los datos se utilizó como soporte tecnológico el software estadístico RStudio 1.2.1335.



## CAPÍTULO III. RESULTADOS Y DISCUSIÓN

### 3.1 Descripción del comportamiento de la serie número de pacientes con cáncer



*Figura 9.* Número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.

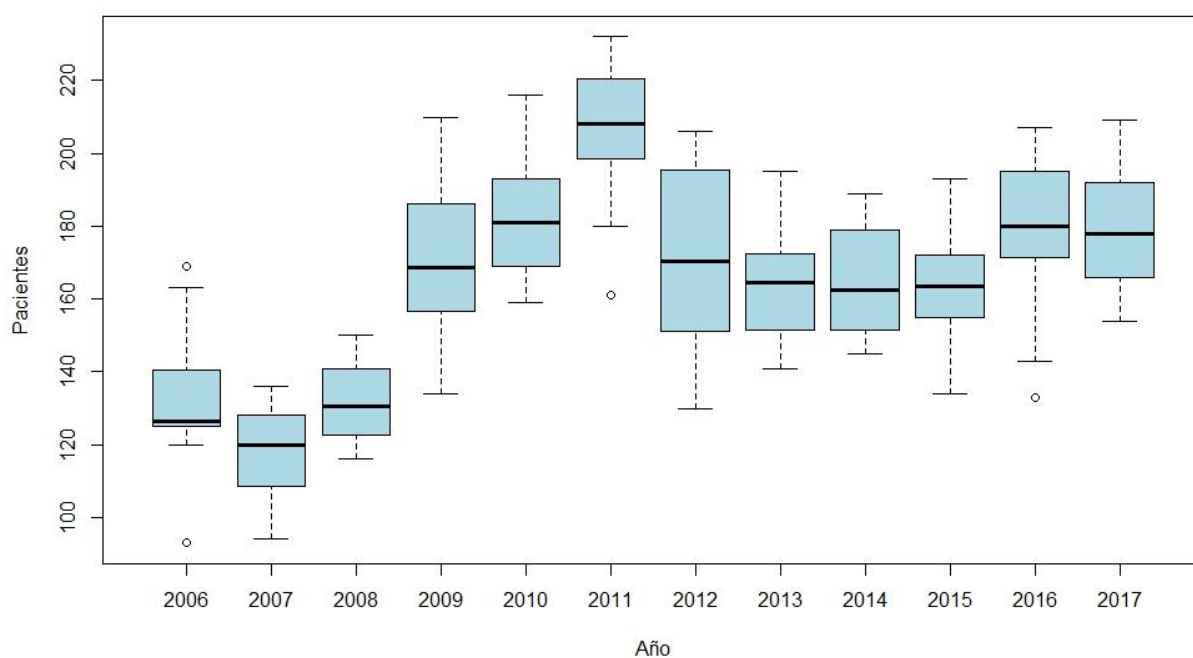
En la figura 9, se observa que en el horizonte del año 2006 hasta el 2017 la serie presenta tendencia creciente y cierta variabilidad a través del tiempo por lo que se puede insinuar que la serie no es estacionaria y requiere ser transformada para estimar un modelo adecuado.

*Tabla 1.* Estadísticos descriptivos del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.

Estadístico	Valor
Media	163
Desviación Estándar	30
Coefficiente de Variación	18.33
Mínimo	93
Máximo	232
Conteo	144

En la tabla 1 se observa, que en promedio se diagnosticaron mensualmente 163 pacientes con cáncer y con una desviación de 30 pacientes, lo que representa una variación porcentual del 18.33%, indicando este valor que los datos registrados mensualmente son homogéneos y por lo tanto presentan una baja variabilidad respecto al promedio.

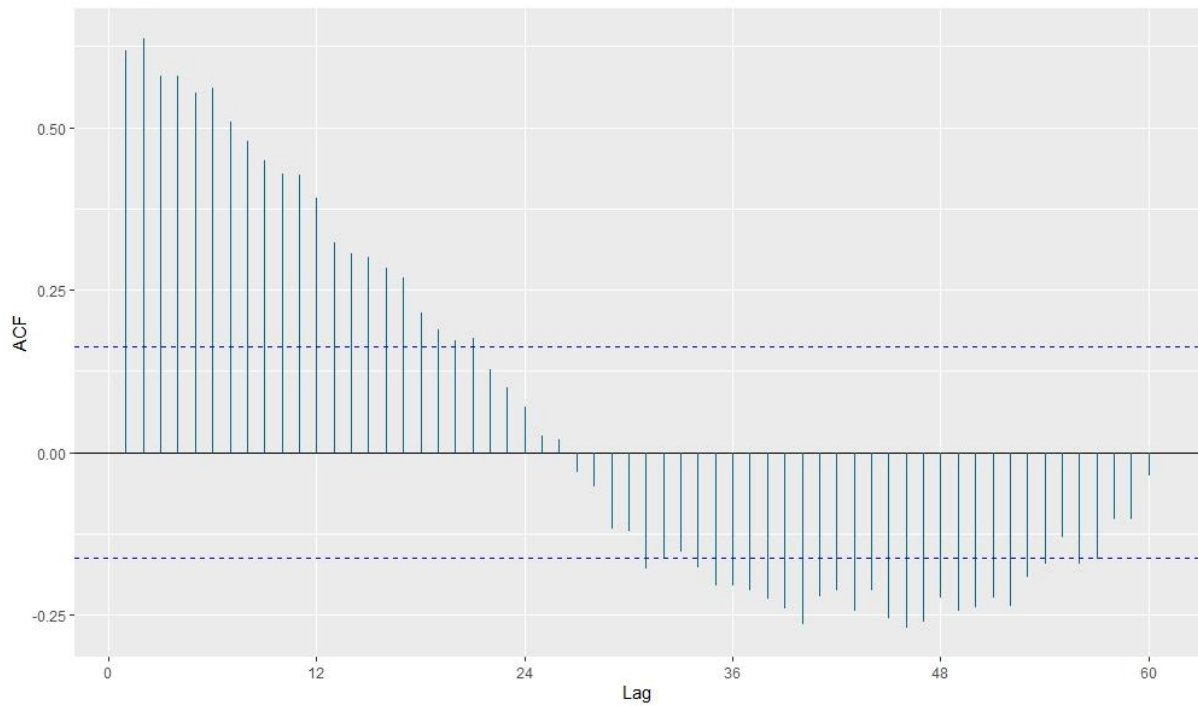
En diciembre del año 2006 se diagnosticaron 93 pacientes con cáncer, siendo este el menor registro mensual y el mayor registro fue de 232 pacientes diagnosticados con cáncer (julio 2011).



*Figura 10.* Diagrama de cajas del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.

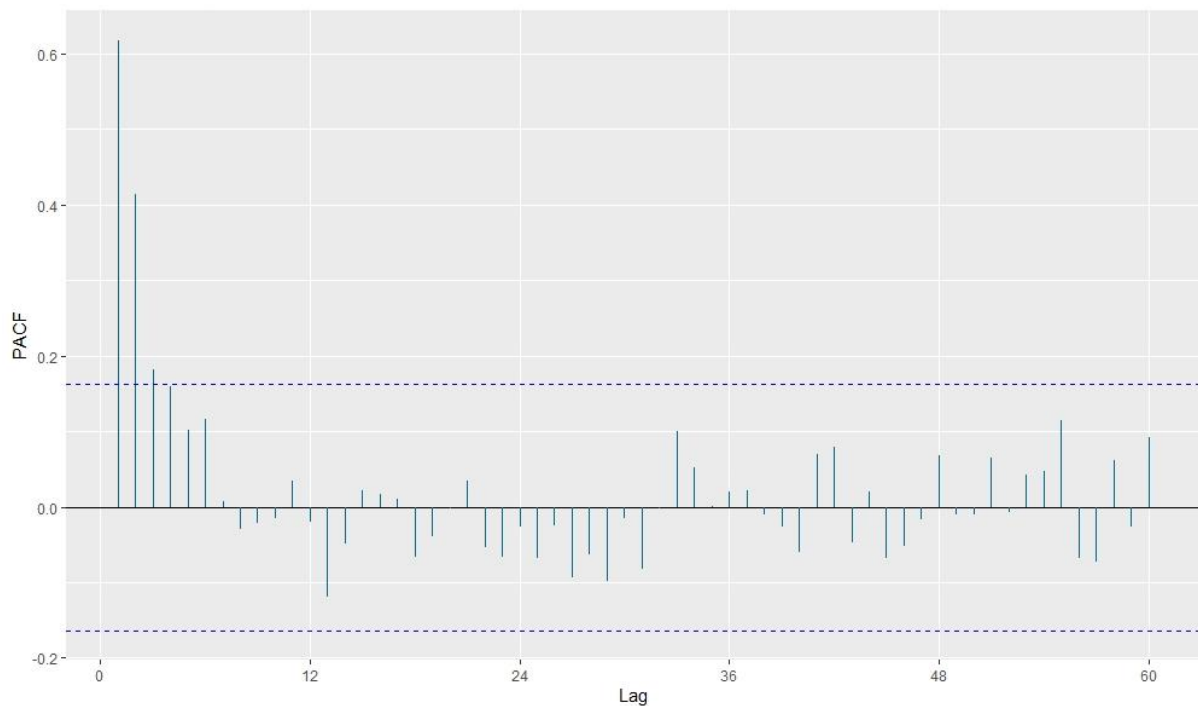
En la figura 10 se observó que en el Hospital Nacional Almanzor Aguinaga Asenjo durante los meses del año 2011 se presentaron los mayores números de pacientes diagnosticados con cáncer, con menor frecuencia se presentaron en los meses de los años 2006, 2007 y 2008, por otro lado en los meses de los años 2009 y 2012 se presentó una alta variabilidad en el número de pacientes diagnosticados con cáncer manifestándose en el ancho de las cajas.

En general, se determinó que el número promedio de pacientes diagnosticados de cáncer por año es variado y tiende a incrementarse, sin embargo, la variabilidad en el número de pacientes diagnosticados de cáncer comparando año a año son similares a excepción de los años 2009 y 2012.



*Figura 11.* Función de autocorrelación simple del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.

La función de autocorrelación simple, es otra forma de representar la serie número de pacientes con cáncer, aquí se puede observar que los coeficientes de autocorrelación decrecen lentamente, al igual que muestra la gráfica de líneas en la cual la tendencia es creciente, por lo que se observa que la serie original no es estacionaria.



*Figura 12.* Función de autocorrelación parcial del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.

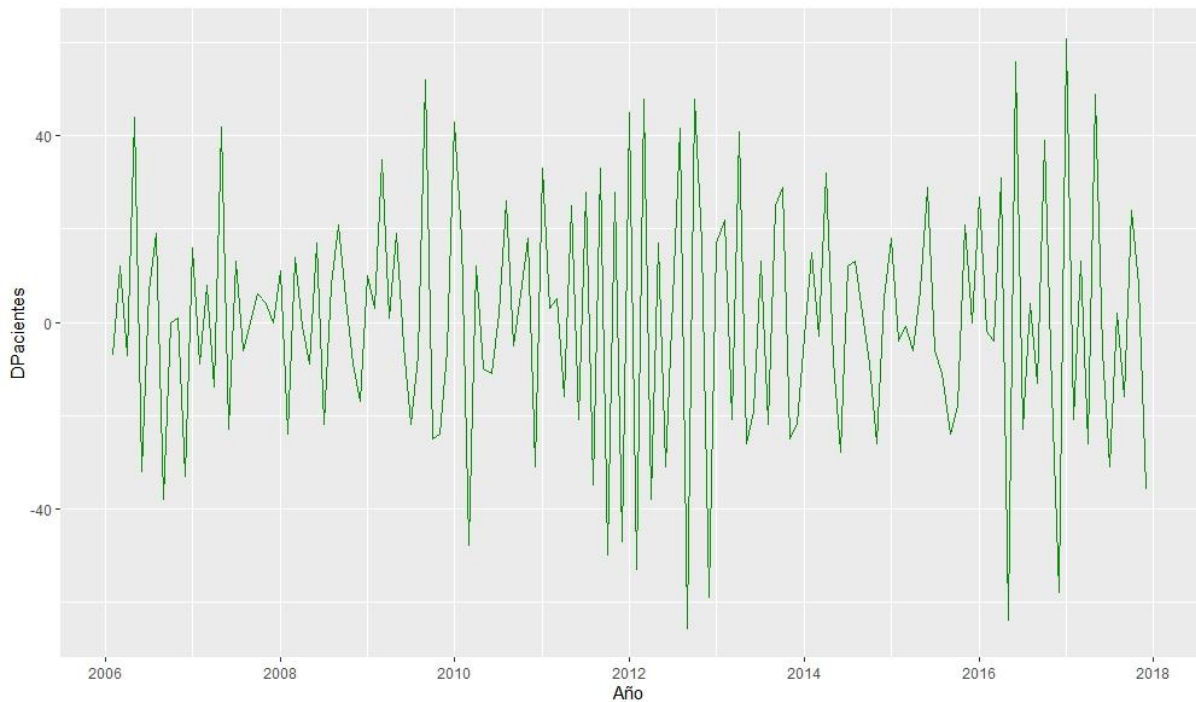
La función de autocorrelación parcial es una herramienta que permite orientar y ayudar a determinar el modelo de serie tiempo. Se aprecia que los coeficientes de autocorrelación descienden rápidamente a cero, pero de manera sinusoidal, además que tres coeficientes son significativos, ya que sobresalen de los límites de confianza, por lo que es posible que exista parte autorregresiva en los modelos de pronósticos buscados para la serie número de pacientes con cáncer en el Hospital Almanzor Aguinaga Asenjo.

Habiendo utilizado 3 herramientas gráficas, se puede deducir que la serie tiene tendencia y estacionalidad, por lo cual se realizaron los test analíticos como se detalla a continuación:

*Tabla 2.* Test de Dickey-Fuller para la serie del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.

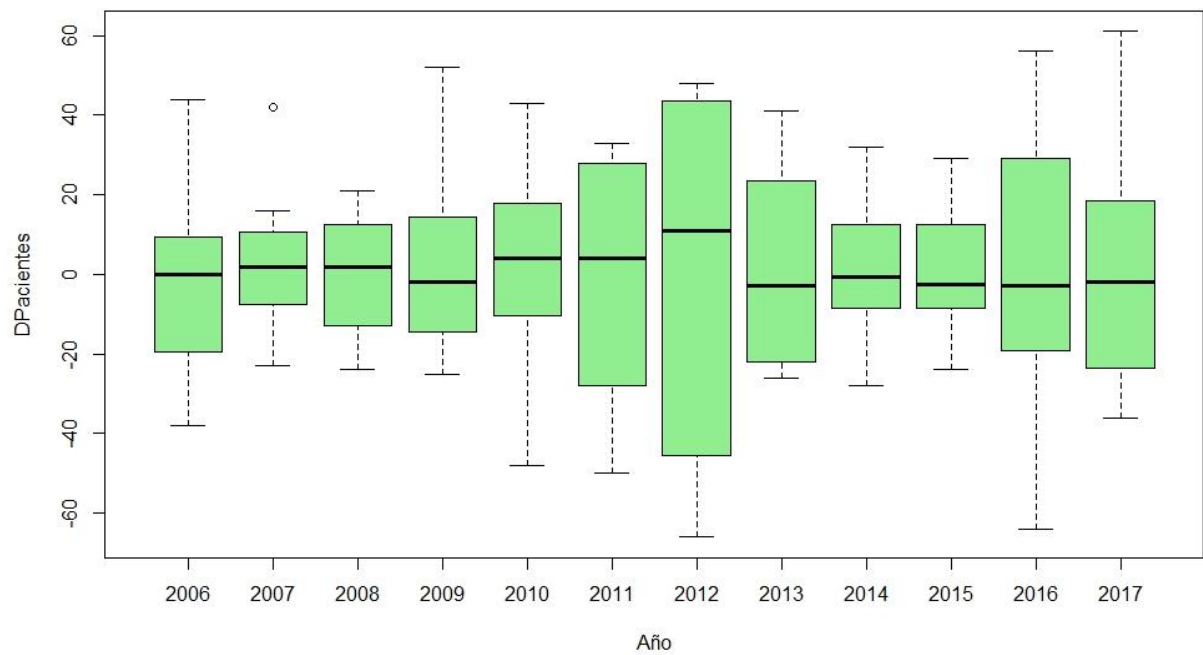
Dickey-Fuller	Lag order	p-value
-1.842	5	0.642

En esta tabla se aprecia el valor del estadístico Dickey – Fuller, el cual es -1.842, lo que permite concluir que no existen evidencias ( $p = 0.642$ ) para desestimar la presencia de raíz unitaria en el proceso generador de los datos de la serie analizada, por lo que varianza de la serie no es estacionaria, por lo tanto para aplicar la metodología de Box y Jenkins se debe realizar una diferencia como se muestra en la siguiente gráfica:



*Figura 13.* Primera diferencia aplicada al número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.

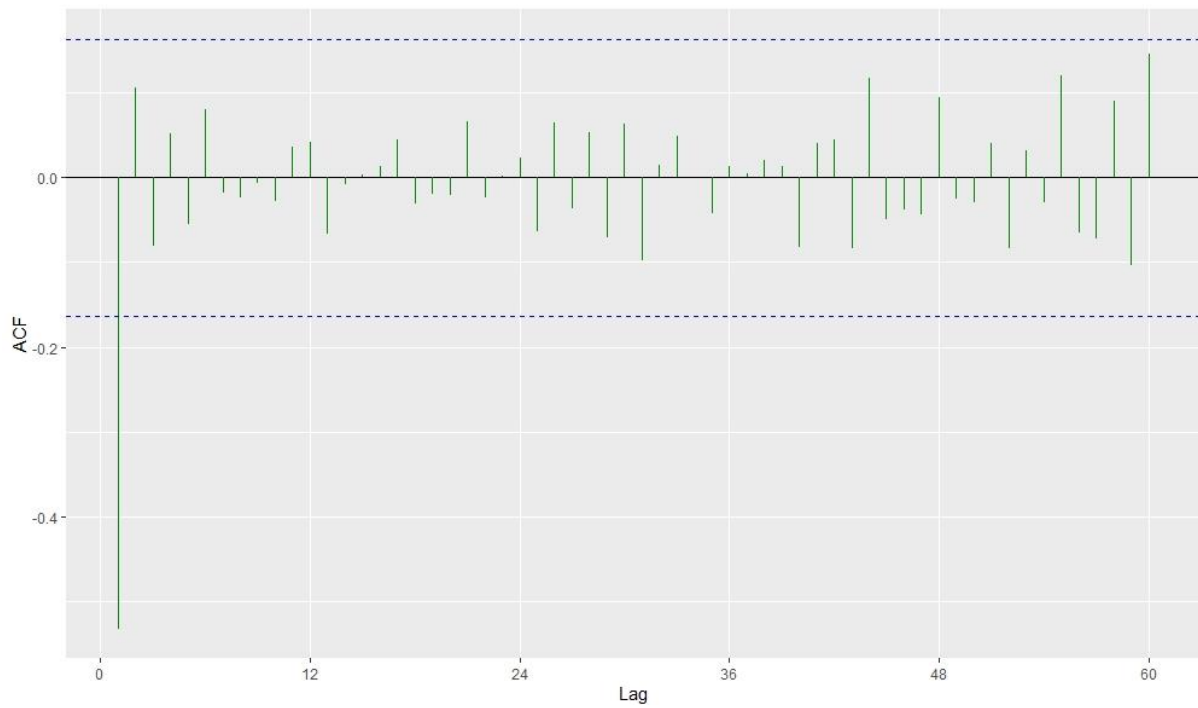
En la figura 13 se presenta la serie diferenciada en un periodo al número de pacientes con cáncer en el Hospital Almanzor Aguinaga Asenjo; se ve que ya desapareció el componente de tendencia, por lo cual se puede observar que la nueva serie se estabiliza respecto a su media, además de presentar varianza constante, por lo tanto se puede deducir que la serie es estacionaria.



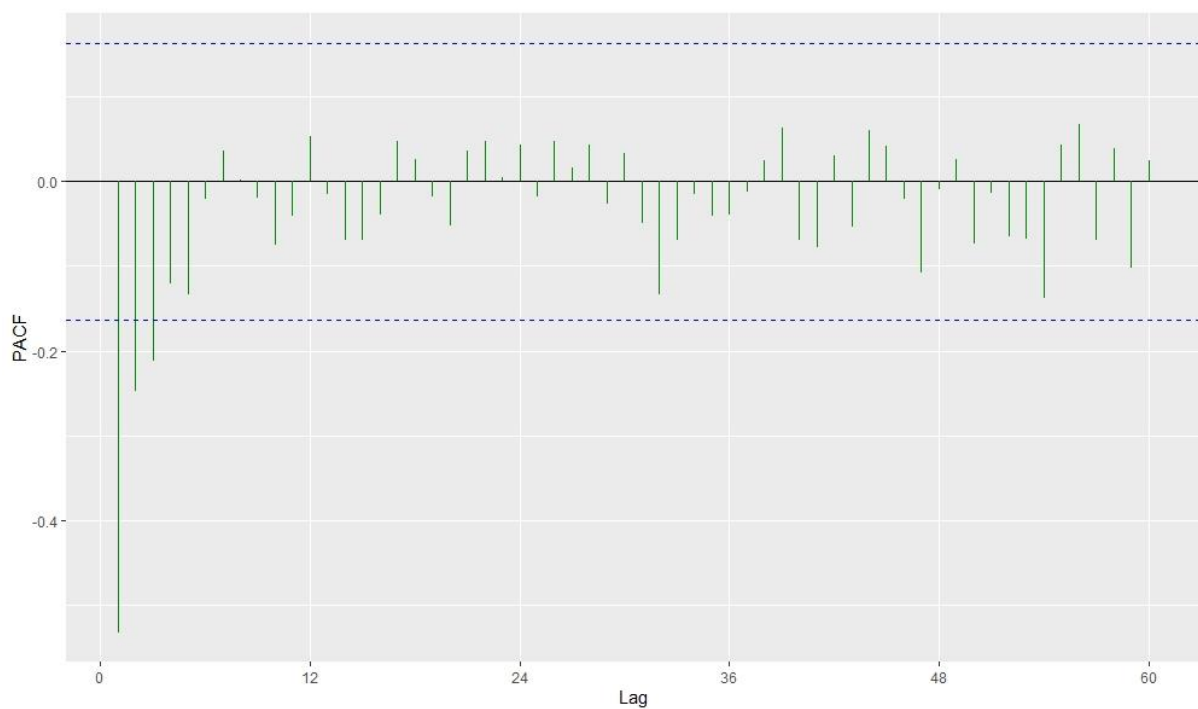
*Figura 14.* Grafica de cajas de la primera diferencia al número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.

Como se observa en la figura, después de aplicar una diferencia en un periodo al número de pacientes con cáncer en el Hospital Nacional Almanzor Aguinaga Asenjo, se puede apreciar que las medianas en las cajas tienden a ser estables; aunque persiste la dispersión entre los pacientes diagnosticados con cáncer en el año 2012. También se observa un dato atípico, por lo cual, se puede decir que la serie ya es estacionaria; lo que se confirmará con la aplicación de la prueba de Dickey Fuller.





*Figura 15.* Función de autocorrelación simple de la primera diferencia del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017



*Figura 16.* Función de autocorrelación parcial de la primera diferencia del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017

En la función de autocorrelación simple de la serie en primera diferencia se observa que los retardos decrecen rápidamente en el tiempo; además se aprecia un retardo significativo, por lo que el modelo propuesto estará formado por un componente MA.

Así mismo, en la función de autocorrelación parcial de la serie en primera diferencia también se aprecia un decrecimiento rápido y la presencia de tres retardos significativos. Este patrón indica un término autorregresivo hasta de orden 3.

*Tabla 3.* Test de Dickey-Fuller de la primera diferencia al número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017.

Dickey-Fuller	Lag order	p-value
-6.782	5	0.01

En esta tabla se aprecia el valor del estadístico Dickey – Fuller, el cual es -6.782, por lo que se concluye que existen evidencias significativas ( $p = 0.01$ ) para rechazar la presencia de raíz unitaria en el modelo, lo que permite establecer que la serie diferenciada es estacionaria.

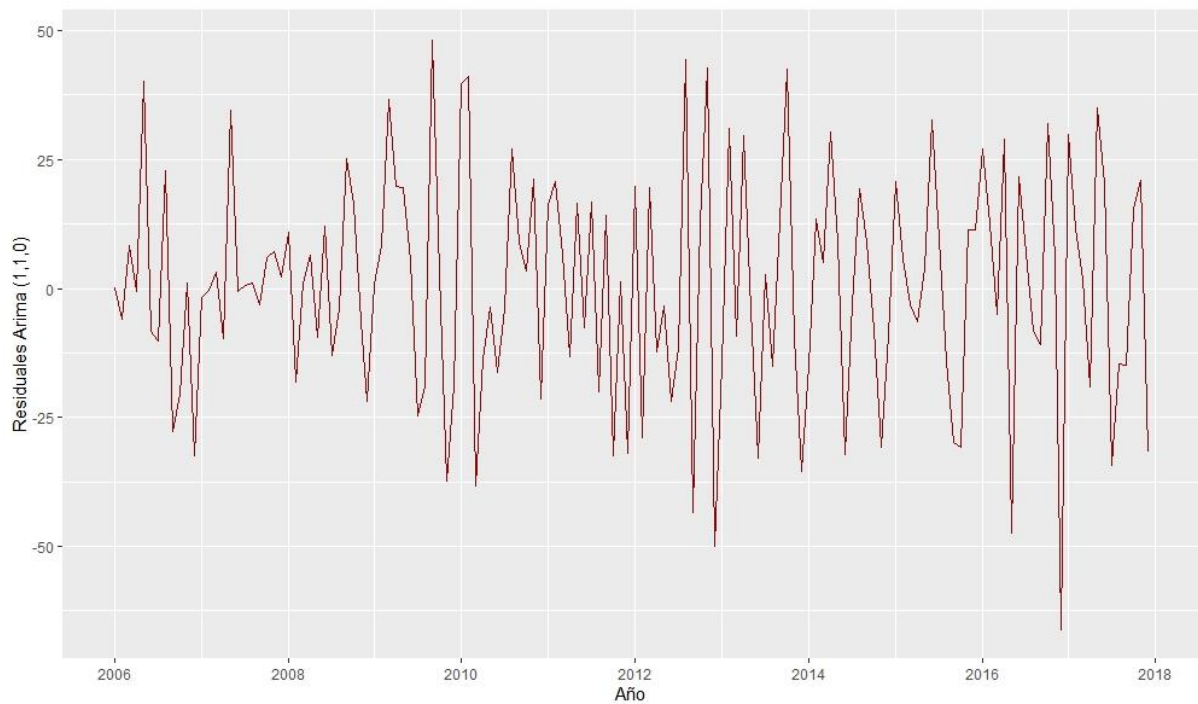
### 3.2 Estimación del modelo de serie de tiempo para el número de pacientes con cáncer, mediante la metodología de Box y Jenkins.

*Tabla 4.* Modelos propuestos que explican el comportamiento de la serie de tiempo número de pacientes con cáncer en el HNAAA

MODELO	PARAMETRO	ESTIMADO	Pr ( $> z $ )
ARIMA (1,1,0)	ar1	-0.536108	3.31E-14
ARIMA (2,1,0)	ar1	-0.671867	2.20E-16
	ar2	-0.251506	0.001894
ARIMA (3,1,0)	ar1	-0.722865	2.20E-16
	ar2	-0.391549	4.87E-05
	ar3	-0.208201	0.01092
SARIMA (1,1,0) (1,1,0)	ar1	-0.523159	1.51E-12
	sar1	-0.468568	1.10E-09
SARIMA (1,1,0) (2,1,0)	ar1	-0.524806	1.32E-12
	sar1	-0.606786	1.14E-12
	sar2	-0.293362	0.0007179
SARIMA (2,1,0) (1,1,0)	ar1	-0.644424	2.92E-14
	ar2	-0.229214	6.60E-03
	sar1	-0.475004	5.79E-10
SARIMA (2,1,0) (2,1,0)	ar1	-0.647204	2.24E-14
	ar2	-0.231854	6.02E-03
	sar1	-0.615373	5.36E-13
	sar2	-0.295434	6.51E-04
SARIMA (3,1,0) (1,1,0)	ar1	-0.697595	2.20E-16
	ar2	-0.37889	1.30E-04
	ar3	-0.228347	7.06E-03
	sar1	-0.49305	8.16E-11
SARIMA (3,1,0) (2,1,0)	ar1	-0.702492	2.20E-16
	ar2	-0.38999	8.50E-05
	ar3	-0.239885	5.01E-03
	sar1	-0.647894	5.86E-14
	sar2	-0.305664	4.44E-04

En la tabla 4 se presentan los diferentes modelos que describen la serie. Posteriormente serán analizados para determinar cuál es el modelo que mejor describe número de pacientes con cáncer en el HNAAA.

## MODELO ARIMA (1,1,0)



*Figura 17.* Residuos del modelo *ARIMA* (1,1,0)

En la figura de los residuos del modelo *ARIMA* (1,1,0), a través del cual se deduce que siguen un ruido blanco.

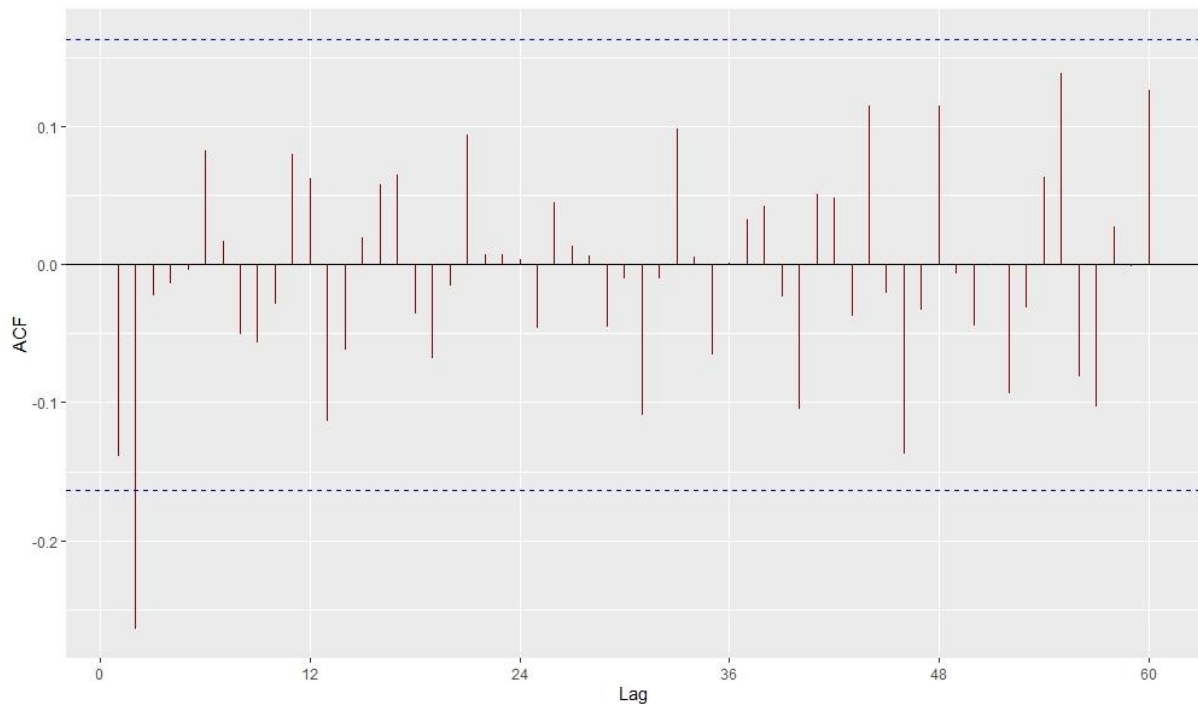


Figura 18. Función de autocorrelación de los residuos del modelo  $ARIMA(1,1,0)$

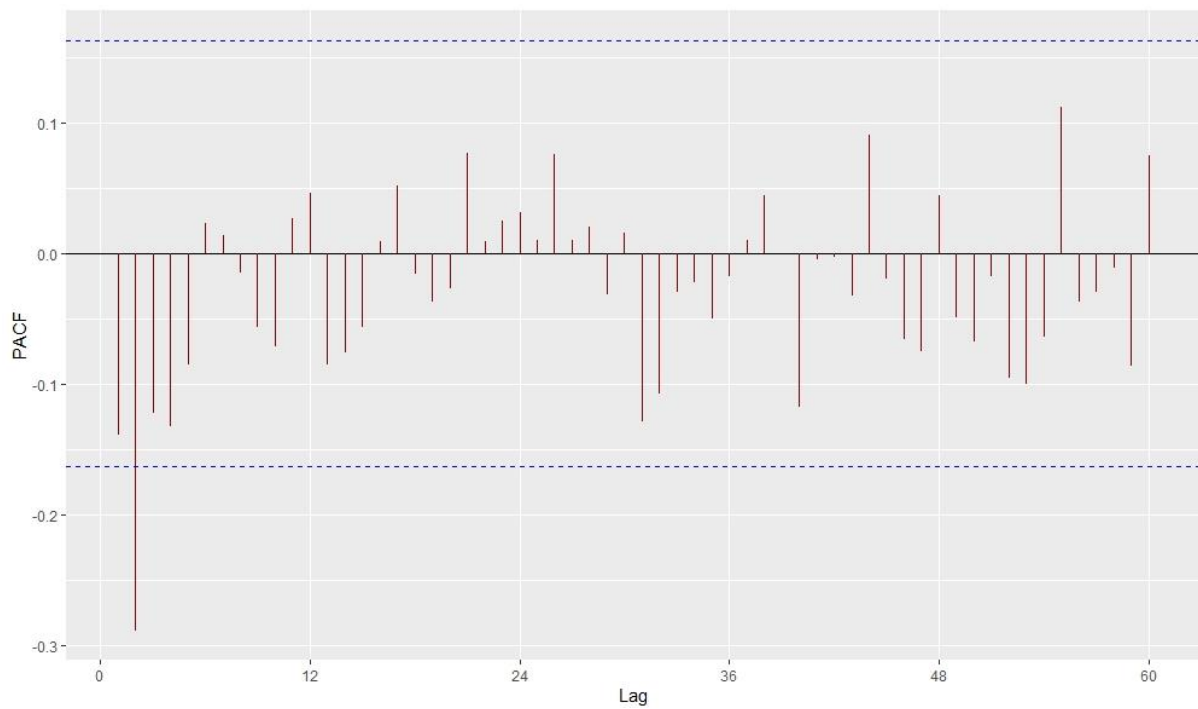


Figura 19. Función de autocorrelación parcial de los residuos del modelo  $ARIMA(1,1,0)$

En las funciones de autocorrelación, tanto simple como parcial de los residuos del modelo *ARIMA* (1,1,0), se observa un retardo que sobresale de los límites de confianza, aun así se deduce que los residuos cumplen con las condiciones de ruido blanco.

*Tabla 5. Test de Ljung-Box de los residuos del modelo ARIMA (1,1,0)*

Chi-Cuadrado	GL	P-Valor
2.828	1	0.093

En la tabla 5 se observa el test de Ljung-Box, el cual arroja un valor chi-cuadrado de 2.828, por lo se determinó que no existen evidencias ( $p = 0.093$ ) para rechazar que todos los coeficientes de autocorrelación son simultáneamente igual a cero, lo que permite concluir que los residuos del modelo *ARIMA* (1,1,0) se distribuyen de forma independiente.

*Tabla 6. Test ANOVA de los residuos del modelo ARIMA (1,1,0)*

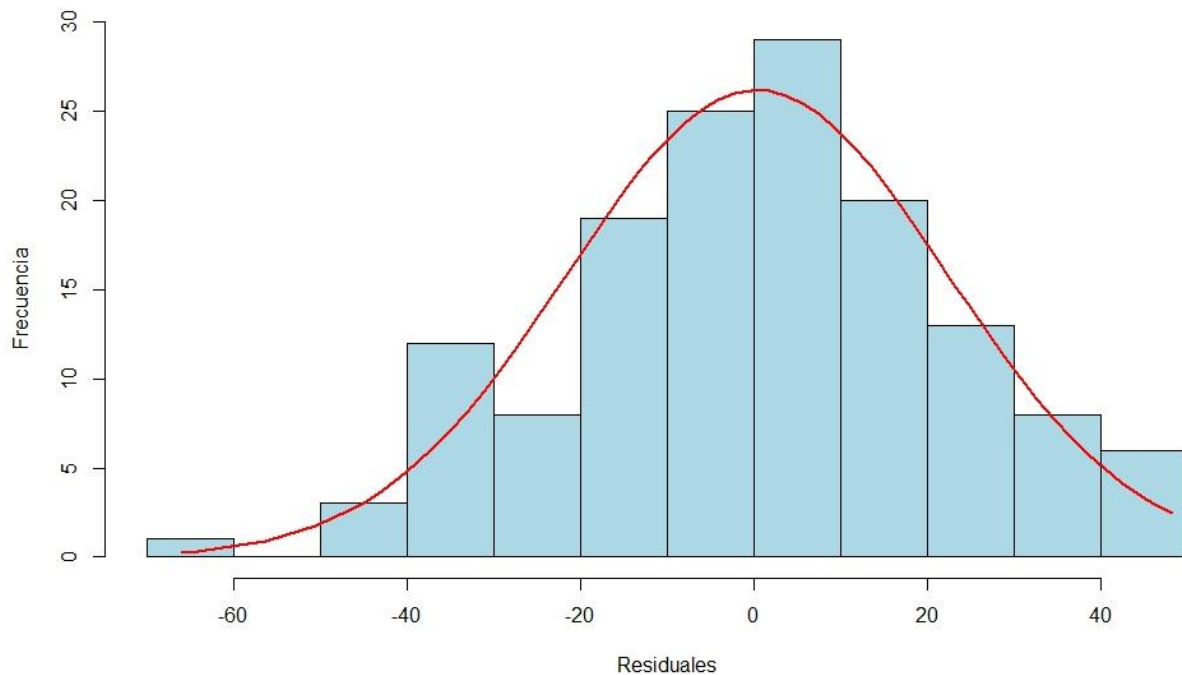
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	3	2.6	0.005	0.942
Residuals	142	68751	484.2		

En la tabla 6 se observa el test ANOVA, el cual arroja un valor F de 0.005, por lo cual se puede determinar que no existen evidencias ( $p = 0.942$ ) para rechazar de que las perturbaciones son variables aleatorias de media nula, por lo que se concluye que los residuos del modelo *ARIMA* (1,1,0) tienen medias igual a cero.

*Tabla 7. Test de Levene de los residuos del modelo ARIMA (1,1,0)*

Df	Df	F value	Pr(>F)
11	132	1.664	0.089

En la tabla 7 se observa el test de Levene, el cual arroja un valor F de 1.664, en consecuencia se determinó que no existen evidencias ( $p = 0.089$ ) para rechazar que todas las perturbaciones tienen las mismas varianzas, por lo que se concluye que los residuos del modelo *ARIMA* (1,1,0) tienen varianzas homogéneas.



*Figura 20. Histograma de los residuos del modelo ARIMA (1,1,0)*

En el histograma se observa que los residuos del modelo *ARIMA* (1,1,0), tienden a seguir una distribución normal.

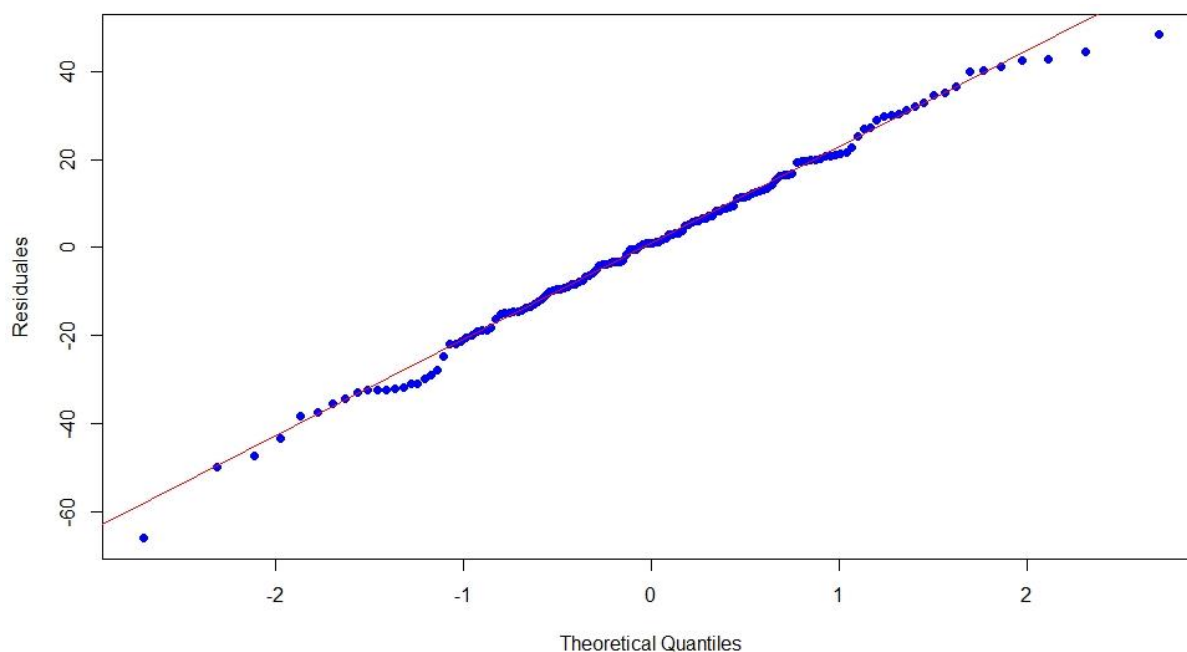


Figura 21. Diagrama Q-Q normal de los residuos del modelo  $ARIMA(1,1,0)$

En el gráfico de probabilidad normal (diagrama Q-Q) muestra que los residuos del modelo  $ARIMA(1,1,0)$ , siguen una línea recta, por lo tanto se deduce que siguen una distribución normal.

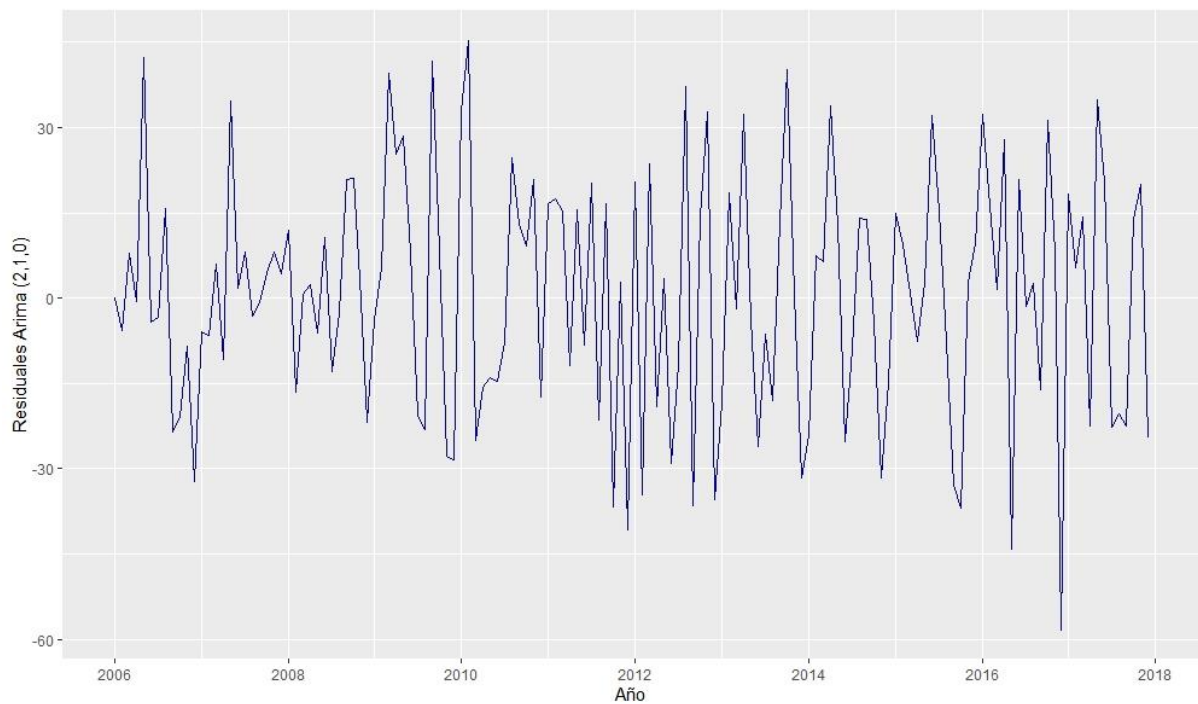
Tabla 8. Test de Kolmogorov-Smirnov de los residuos del modelo  $ARIMA(1,1,0)$

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.035	0.995

En la tabla 8 se aprecia el estadístico de Kolmogorov - Smirnov, la cual arroja un valor de 0.035, por lo que se determinó que no existen evidencias ( $p = 0.995$ ) para rechazar la hipótesis de que las perturbaciones tienen una distribución conjunta normal, concluyendo que los residuos del modelo  $ARIMA(1,1,0)$  siguen una distribución normal.



## MODELO ARIMA (2,1,0)



*Figura 22. Residuos del Pronóstico del modelo ARIMA (2,1,0)*

En la figura de los residuos del modelo *ARIMA* (2,1,0), a través del cual se deduce que siguen un ruido blanco.

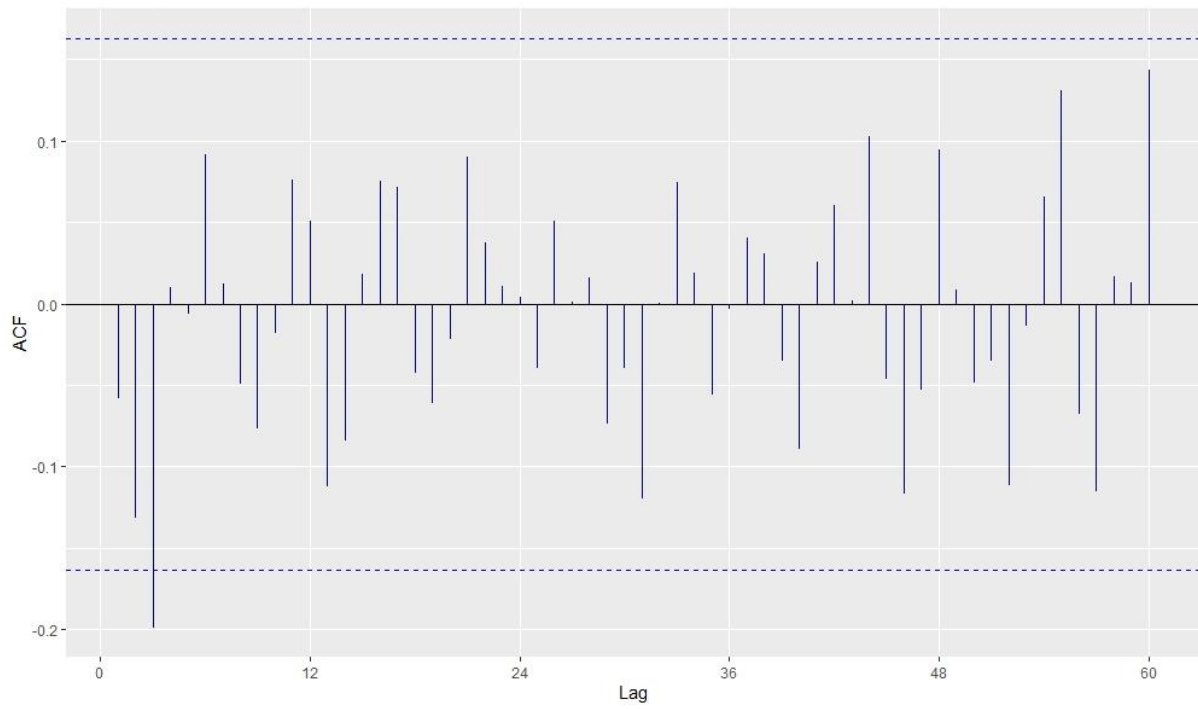


Figura 23. Función de autocorrelación de los residuos del modelo  $ARIMA(2,1,0)$

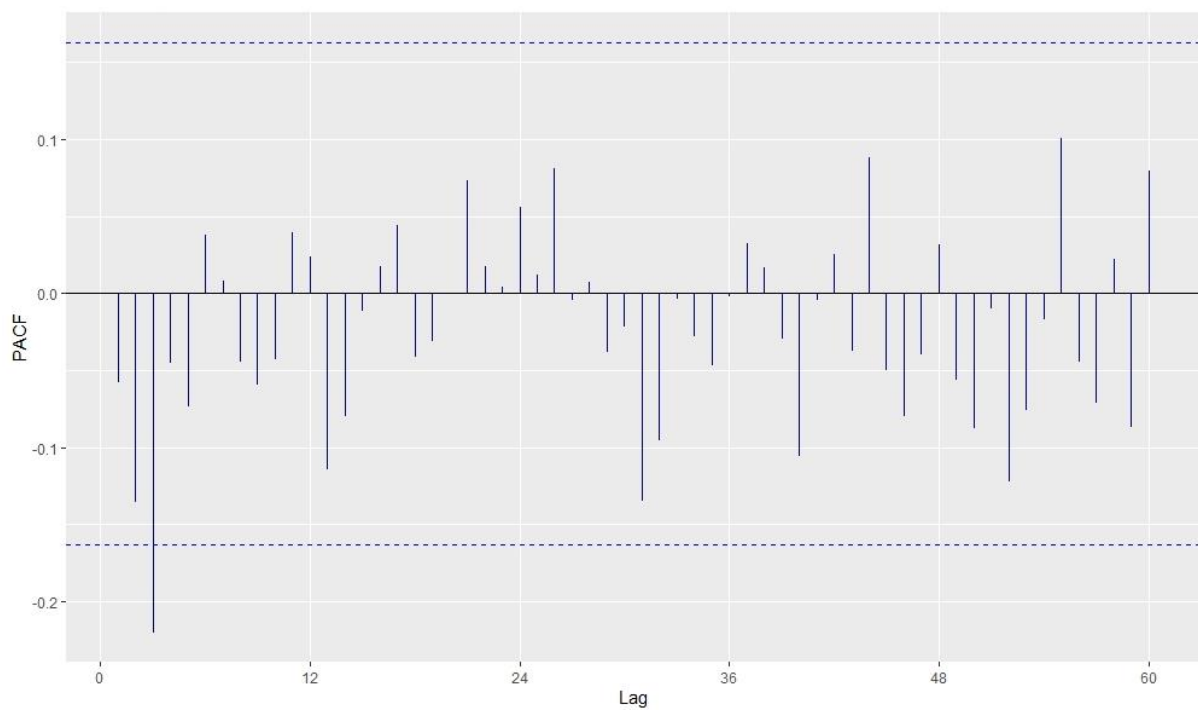


Figura 24. Función de autocorrelación parcial de los residuos del modelo  $ARIMA(2,1,0)$

En las funciones de autocorrelación, tanto simple como parcial de los residuos del modelo *ARIMA* (2,1,0), se observa un retardo que sobresale de los límites de confianza, aun así se deduce que los residuos cumplen con las condiciones de ruido blanco.

*Tabla 9.* Test de Ljung-Box de los residuos del modelo *ARIMA* (2,1,0)

Chi-Cuadrado	GL	P-Valor
0.492	1	0.483

En la tabla 9 se observa el test de Ljung-Box, el cual arroja un valor chi-cuadrado de 0.492, por lo se determinó que no existen evidencias ( $p = 0.483$ ) para rechazar que todos los coeficientes de autocorrelación son simultáneamente igual a cero, lo que permite concluir que los residuos del modelo *ARIMA* (2,1,0) se distribuyen de forma independiente.

*Tabla 10.* Test ANOVA de los residuos del modelo *ARIMA* (2,1,0)

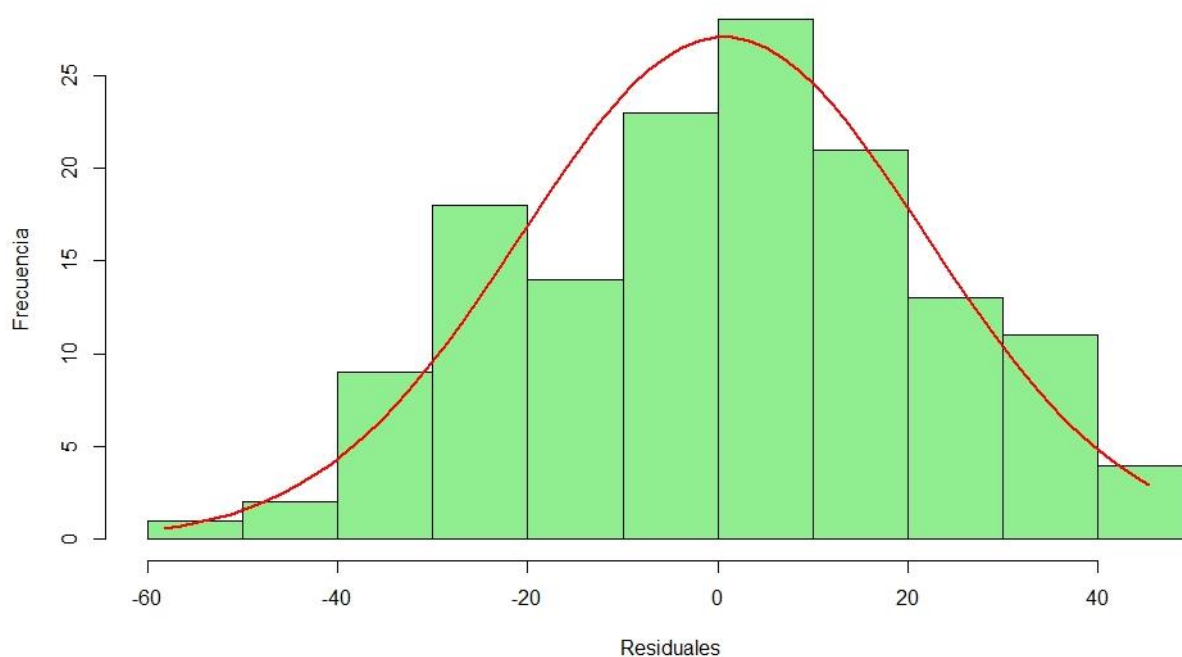
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	22	22.1	0.051	0.821
Residuals	142	61443	432.7		

En la tabla 10 se observa el test ANOVA, el cual arroja un valor F de 0.051, por lo cual se puede determinar que no existen evidencias ( $p = 0.821$ ) para rechazar de que las perturbaciones son variables aleatorias de media nula, por lo que se concluye que los residuos del modelo *ARIMA* (2,1,0) tienen medias igual a cero.

*Tabla 11. Test de Levene de los residuos del modelo ARIMA (2,1,0)*

Df	Df	F value	Pr(>F)
11	132	1.713	0.077

En la tabla 11 se observa el test de Levene, el cual arroja un valor F de 1.713, en consecuencia se determinó que no existen evidencias ( $p = 0.077$ ) para rechazar que todas las perturbaciones tienen las mismas varianzas, por lo que se concluye que los residuos del modelo *ARIMA* (2,1,0) tienen varianzas homogéneas.



*Figura 25. Histograma de los residuos del modelo ARIMA (2,1,0)*

En el histograma se observa que los residuos del modelo *ARIMA* (2,1,0), tienden a seguir una distribución normal.

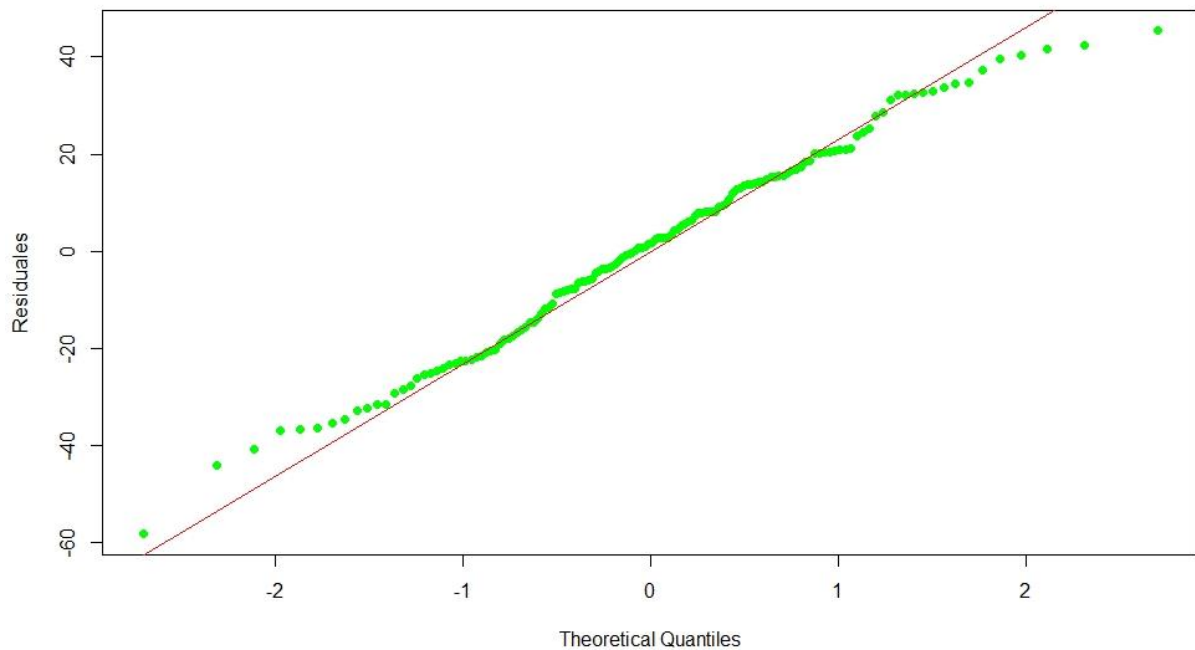


Figura 26. Diagrama Q-Q normal de los residuos del modelo  $ARIMA(2,1,0)$

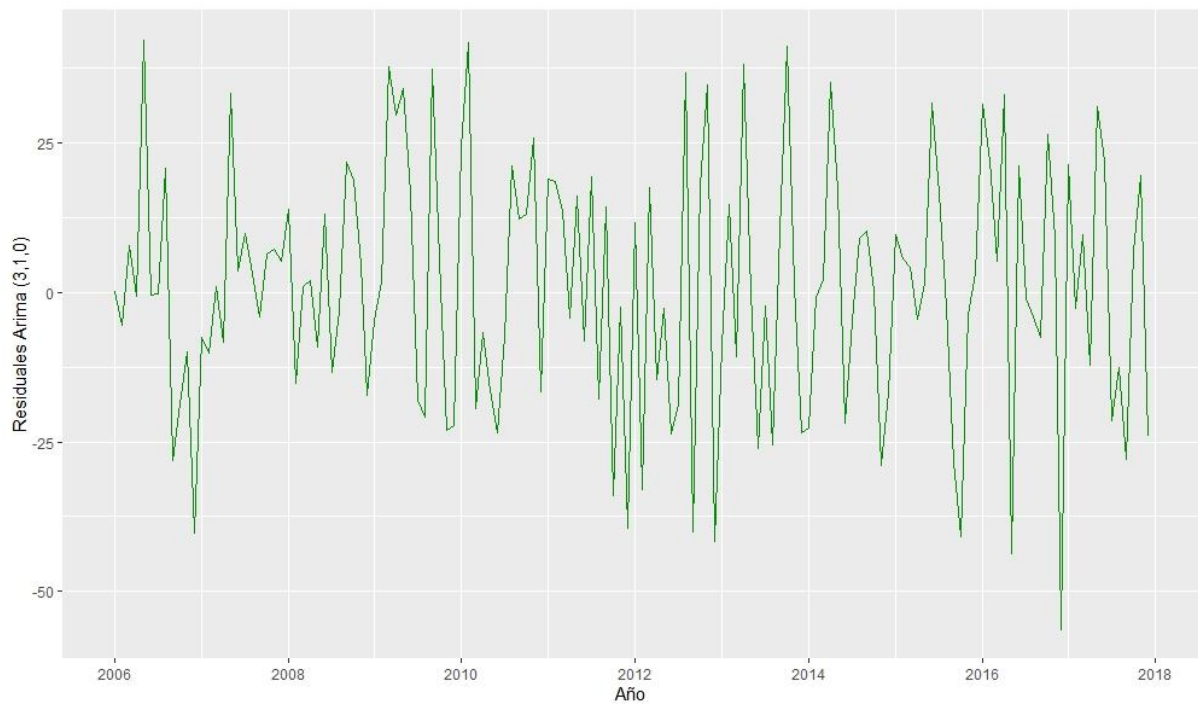
En el gráfico de probabilidad normal (diagrama Q-Q) muestra que los residuos del modelo  $ARIMA(2,1,0)$ , siguen una línea recta, por lo tanto se deduce que siguen una distribución normal.

Tabla 12. Test de Kolmogorov-Smirnov de los residuos del modelo  $ARIMA(2,1,0)$

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.046	0.915

En la tabla 12 se aprecia el estadístico de Kolmogorov - Smirnov, la cual arroja un valor de 0.046, por lo que se determinó que no existen evidencias ( $p = 0.915$ ) para rechazar la hipótesis de que las perturbaciones tienen una distribución conjunta normal, concluyendo que los residuos del modelo  $ARIMA(2,1,0)$  siguen una distribución normal.

## MODELO ARIMA (3,1,0)



*Figura 27. Residuos del modelo ARIMA (3,1,0)*

En la figura 27 se observan los residuos del modelo *ARIMA* (3,1,0), a través de la cual se deduce que siguen un ruido blanco.



Figura 28. Función de autocorrelación de los residuos del modelo  $ARIMA(3,1,0)$



Figura 29. Función de autocorrelación parcial de los residuos del modelo  $ARIMA(3,1,0)$

En las funciones de autocorrelación, tanto simple como parcial de los residuos del modelo *ARIMA* (3,1,0), se observa que los retardos están dentro de los límites de confianza, por lo que se deduce que los residuos cumplen con las condiciones de ruido blanco.

*Tabla 13.* Test de Ljung-Box de los residuos del modelo *ARIMA* (3,1,0)

Chi-Cuadrado	GL	P-Valor
0.145	1	0.704

En la tabla 13 se observa el test de Ljung-Box, el cual arroja un valor chi-cuadrado de 0.145, por lo que se determinó que no existen evidencias ( $p = 0.704$ ) para rechazar que todos los coeficientes de autocorrelación son simultáneamente igual a cero, lo que permite concluir que los residuos del modelo *ARIMA* (3,1,0) se distribuyen de forma independiente.

*Tabla 14.* Test ANOVA de los residuos del modelo *ARIMA* (3,1,0)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	11	11.1	0.017	0.896
Residuals	142	90981	640.7		

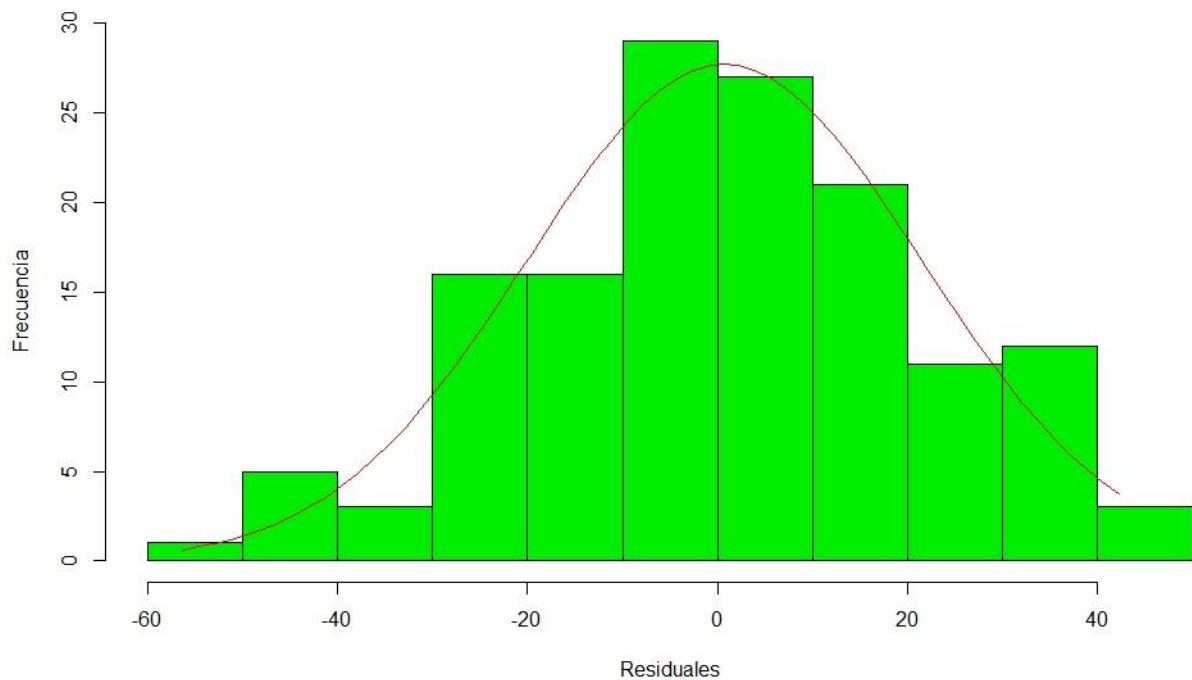
En la tabla 14 se observa el test ANOVA, el cual arroja un valor F de 0.017, por lo que se puede determinar que no existen evidencias ( $p = 0.896$ ) para rechazar de que las perturbaciones son variables aleatorias de media nula, por lo que se concluye que los residuos del modelo *ARIMA* (3,1,0) tienen medias igual a cero.



*Tabla 15. Test de Levene de los residuos del modelo ARIMA (3,1,0)*

Df	Df	F value	Pr(>F)
11	132	1.669	0.087

En la tabla 15 se observa el test de Levene, el cual arroja un valor F de 1.669, en consecuencia se determinó que no existen evidencias ( $p = 0.087$ ) para rechazar que todas las perturbaciones tienen las mismas varianzas, por lo que se concluye que los residuos del modelo *ARIMA* (3,1,0) tienen varianzas homogéneas.



*Figura 30. Histograma de los residuos del Pronóstico del modelo ARIMA (3,1,0)*

En el histograma se observa que los residuos del modelo *ARIMA* (3,1,0), tienden a seguir una distribución normal.

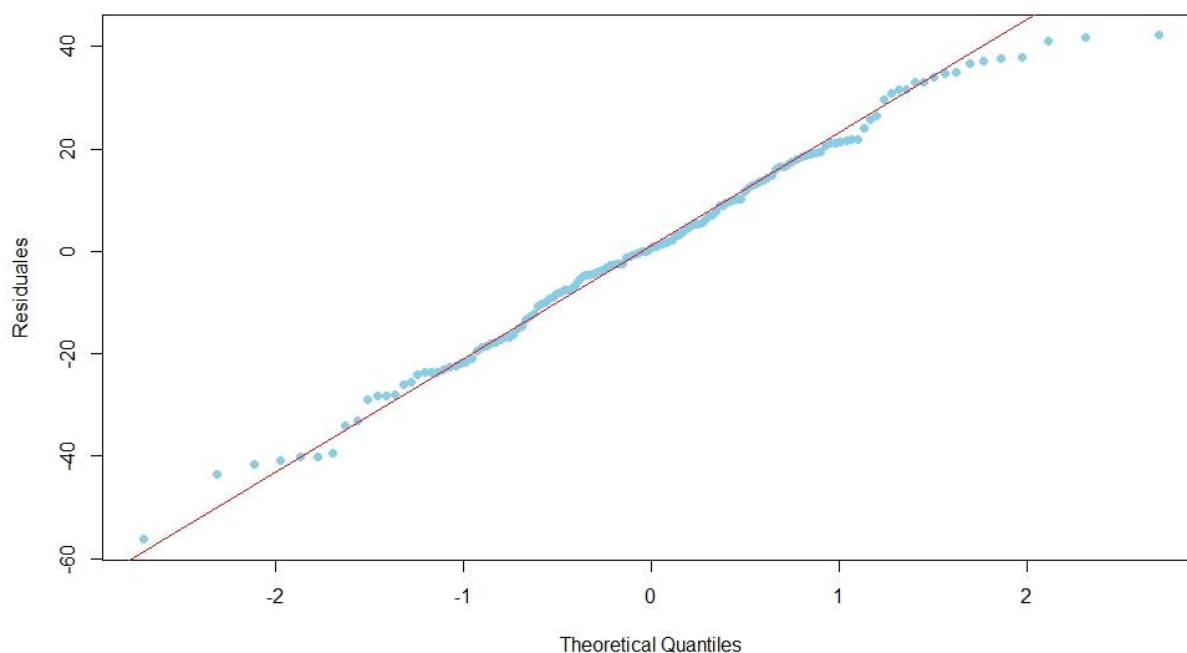


Figura 31. Diagrama Q-Q normal de los residuos del modelo  $ARIMA(3,1,0)$

En el gráfico de probabilidad normal (diagrama Q-Q) muestra que los residuos del modelo  $ARIMA(3,1,0)$ , siguen una línea recta, por lo tanto se deduce que siguen una distribución normal.

Tabla 16. Test de Kolmogorov-Smirnov de los residuos del modelo  $ARIMA(3,1,0)$

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.041	0.966

En la tabla 16 se aprecia el estadístico de Kolmogorov - Smirnov, la cual arroja un valor de 0.041, por lo que se determinó que no existen evidencias ( $p = 0.966$ ) para rechazar la hipótesis de que las perturbaciones tienen una distribución conjunta normal, concluyendo que los residuos del modelo  $ARIMA(3,1,0)$  siguen una distribución normal.

### MODELO SARIMA (1,1,0) (1,1,0)<sub>12</sub>

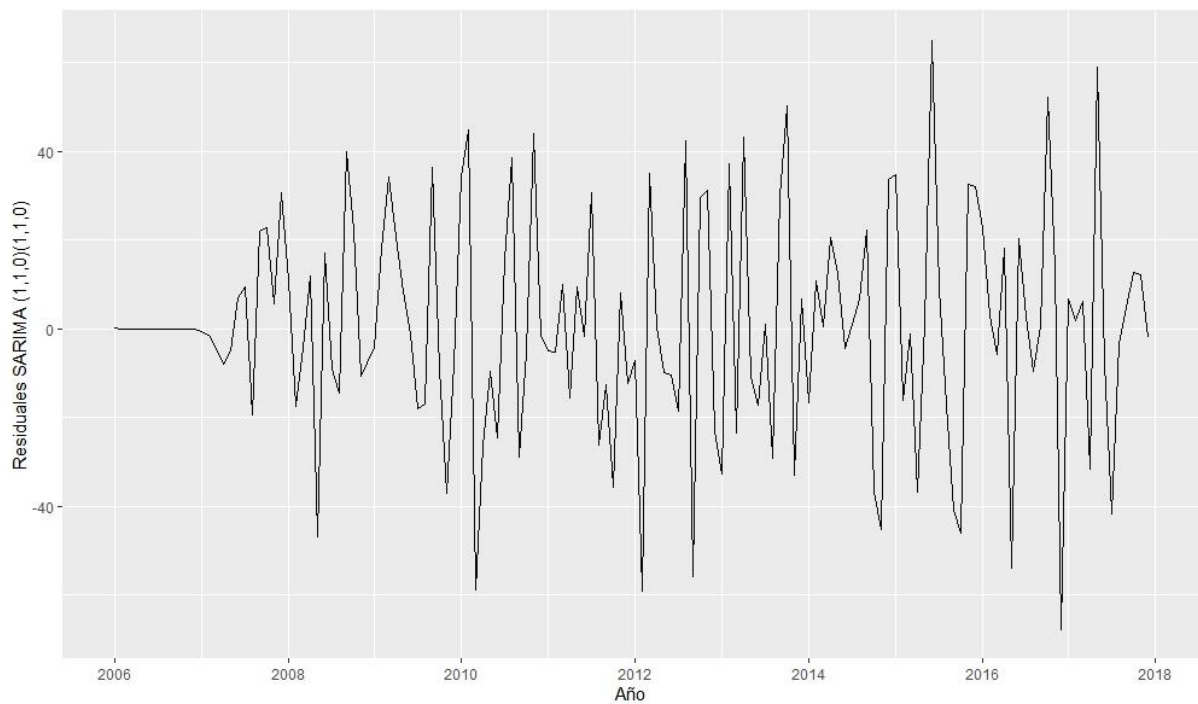


Figura 32. Residuos del modelo  $SARIMA (1,1,0)(1,1,0)_{12}$

En la figura 32 se presentan los residuos del modelo  $SARIMA (1,1,0)(1,1,0)_{12}$ , a través de la cual se puede observar que siguen un ruido blanco.

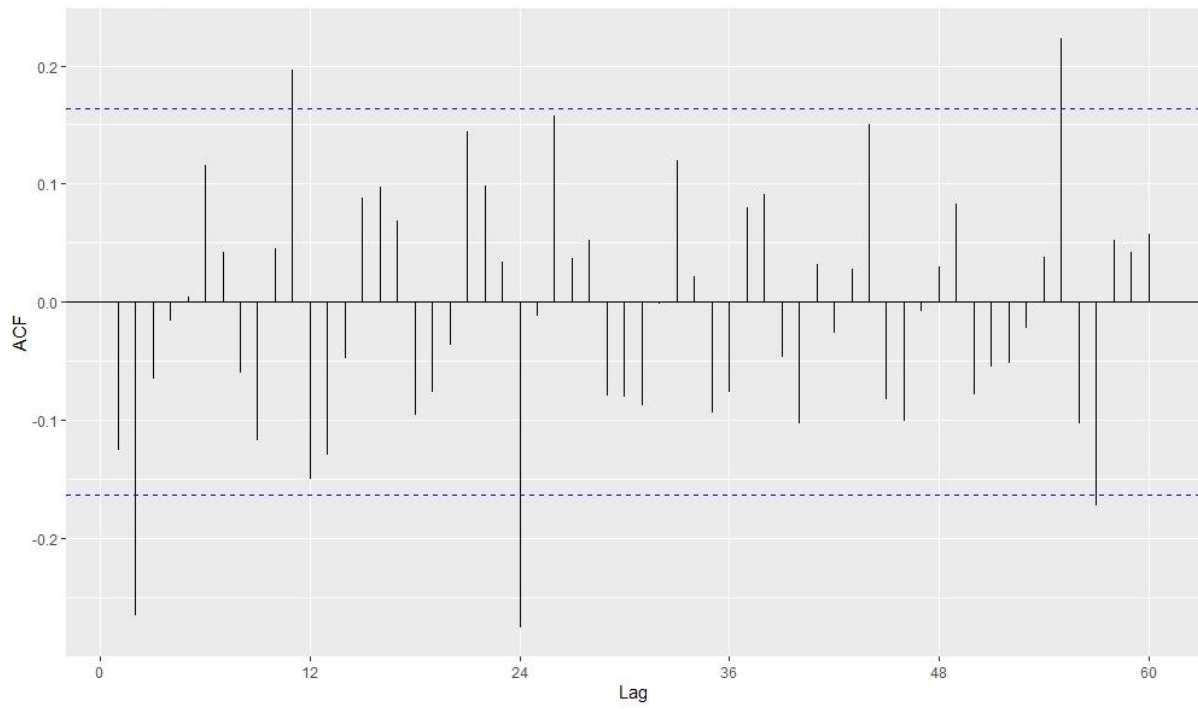


Figura 33. Función de autocorrelación de los residuos del modelo  $SARIMA (1,1,0)(1,1,0)_{12}$

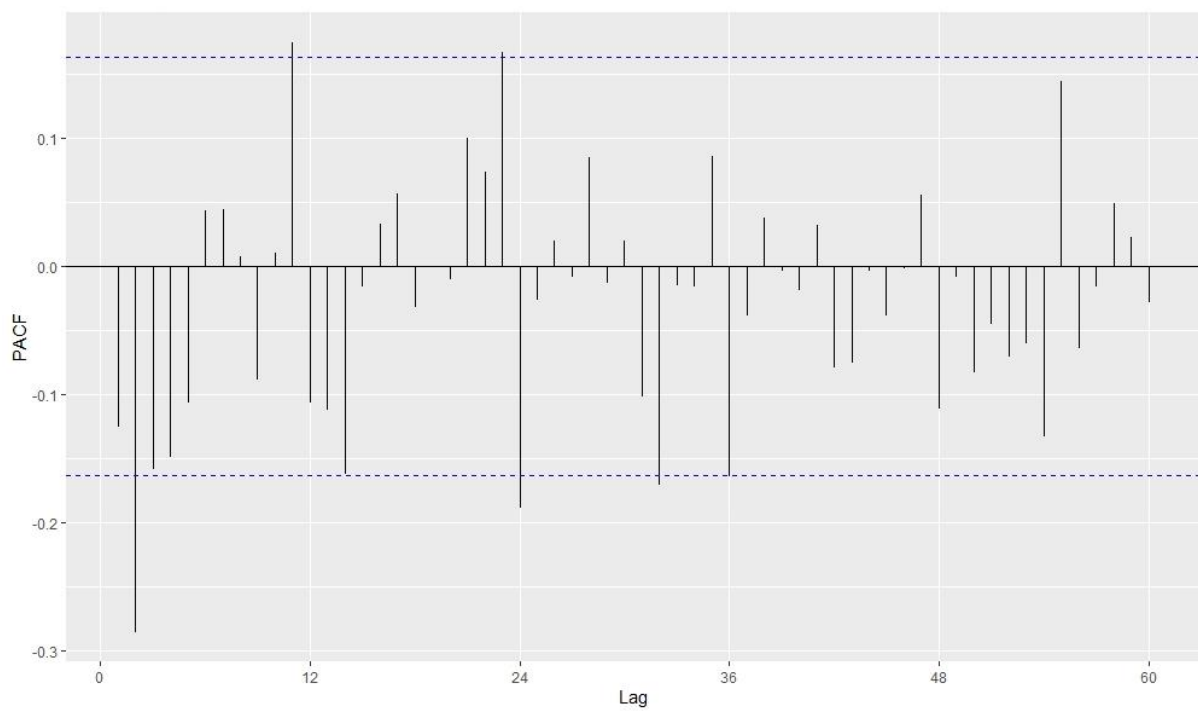


Figura 34. Función de autocorrelación parcial de los residuos del modelo  $SARIMA (1,1,0)(1,1,0)_{12}$

En las funciones de autocorrelación, tanto simple como parcial de los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$ , se observa que existen retardos que sobresalen de los límites de confianza, por lo que se deduce que los residuos no cumplirían con las condiciones de ruido blanco.

*Tabla 17.* Test de Ljung-Box de los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$

Chi-Cuadrado	GL	P-Valor
2.310	1	0.129

En la tabla 17 se observa el test de Ljung-Box, el cual arroja un valor chi-cuadrado de 2.310, por lo se determinó que no existen evidencias ( $p = 0.129$ ) para rechazar que todos los coeficientes de autocorrelación son simultáneamente igual a cero, lo que permite concluir que los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  se distribuyen de forma independiente.

*Tabla 18.* Test ANOVA de los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$

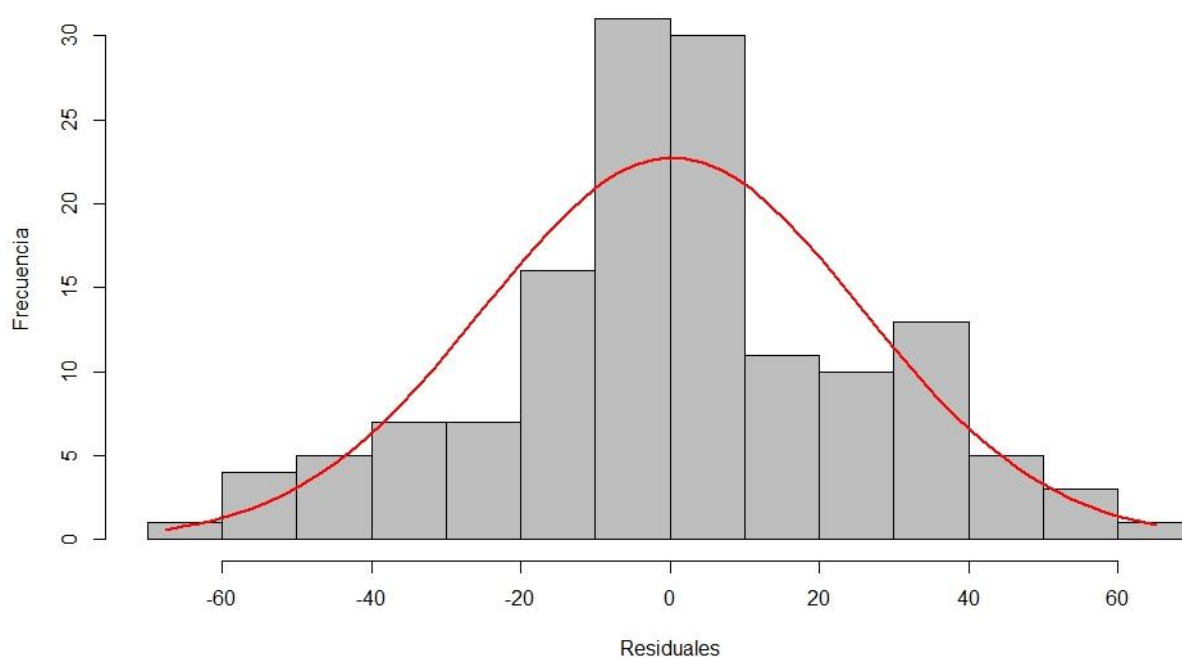
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	11	11.1	0.017	0.896
Residuals	142	90981	640.7		

En la tabla 18 se observa el test ANOVA, el cual arroja un valor F de 0.017, por lo cual se puede determinar que no existen evidencias ( $p = 0.896$ ) para rechazar de que las perturbaciones son variables aleatorias de media nula, por lo que se concluye que los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  tienen medias igual a cero.

*Tabla 19.* Test de Levene de los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$

Df	Df	F value	Pr(>F)
11	132	2.977	0.001

En la tabla 19 se observa el test de Levene, el cual arroja un valor F de 2.977, en consecuencia se determinó que existen evidencias ( $p = 0.001$ ) para rechazar que todas las perturbaciones tienen las mismas varianzas, por lo que se concluye que los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  no tienen varianzas homogéneas.



*Figura 35.* Histograma de los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$

En el histograma se observa que los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$ , tienden a seguir una distribución normal.

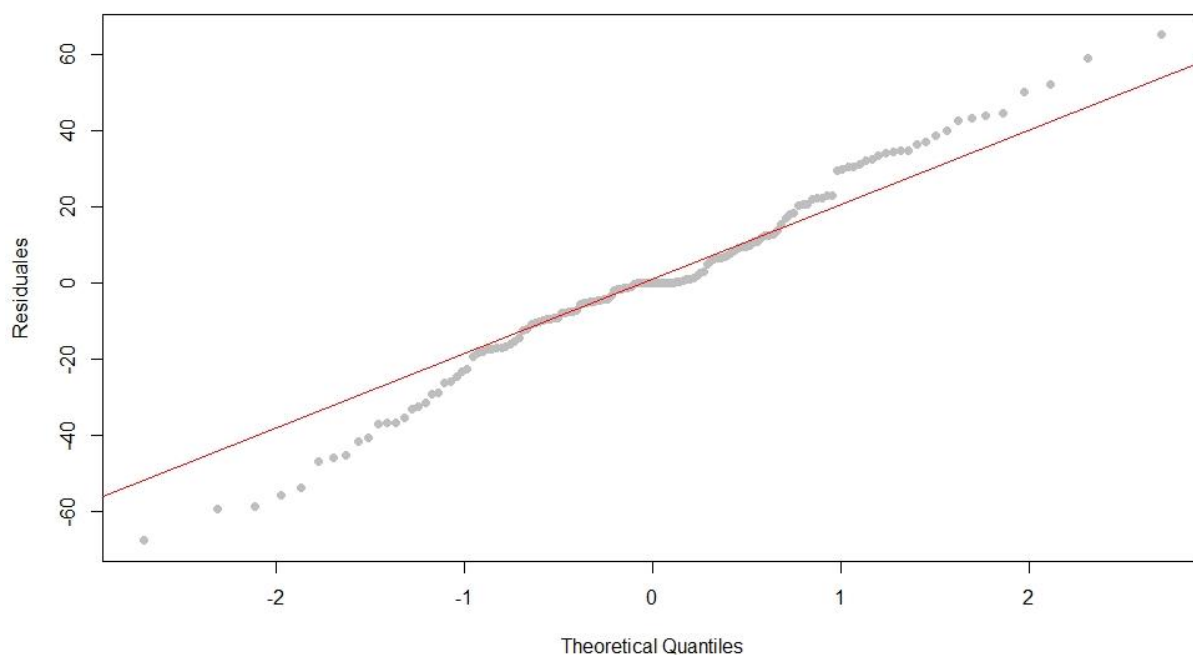


Figura 36. Diagrama Q-Q normal de los residuos del modelo  $SARIMA (1,1,0)(1,1,0)_{12}$

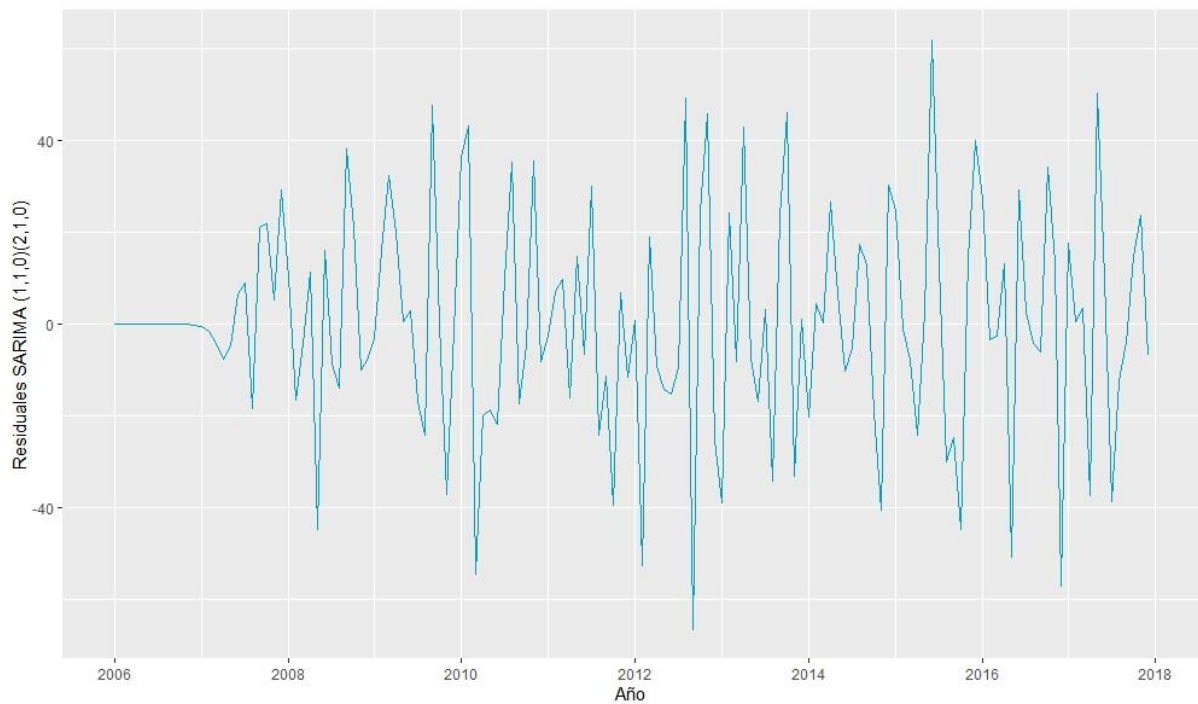
En el gráfico de probabilidad normal (diagrama Q-Q) muestra que los residuos del modelo  $SARIMA (1,1,0)(1,1,0)_{12}$ , siguen una línea recta, por lo tanto se deduce que siguen una distribución normal.

Tabla 20. Test de Kolmogorov-Smirnov de los residuos del modelo  $SARIMA (1,1,0)(1,1,0)_{12}$

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.074	0.406

En la tabla 20 se aprecia el estadístico de Kolmogorov - Smirnov, la cual arroja un valor de 0.074, por lo que se determinó que no existen evidencias ( $p = 0.406$ ) para rechazar la hipótesis de que las perturbaciones tienen una distribución conjunta normal, concluyendo que los residuos del modelo  $SARIMA (1,1,0)(1,1,0)_{12}$  siguen una distribución normal.

### MODELO SARIMA (1,1,0) (2,1,0)<sub>12</sub>



*Figura 37.* Residuos del modelo  $SARIMA (1,1,0)(2,1,0)_{12}$

En la figura 37 se tienen los residuos del modelo  $SARIMA (1,1,0)(2,1,0)_{12}$ , por la cual se observa que siguen un ruido blanco.



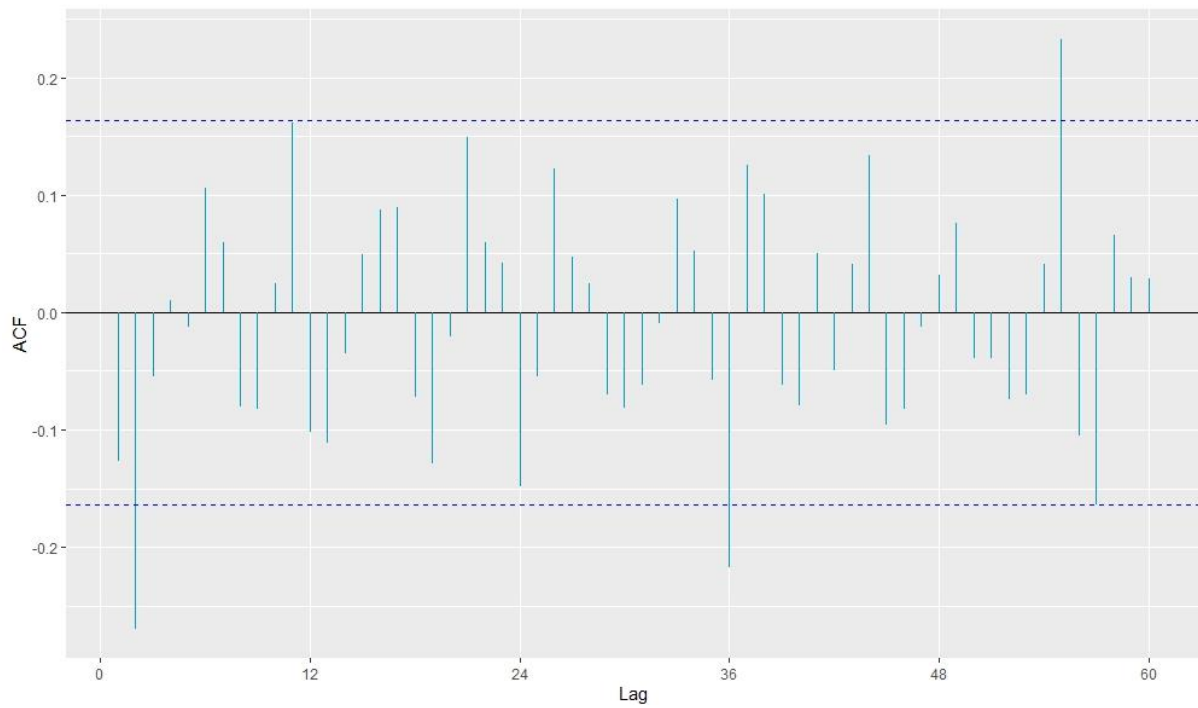


Figura 38. Función de autocorrelación de los residuos del modelo  $SARIMA (1,1,0)(2,1,0)_{12}$

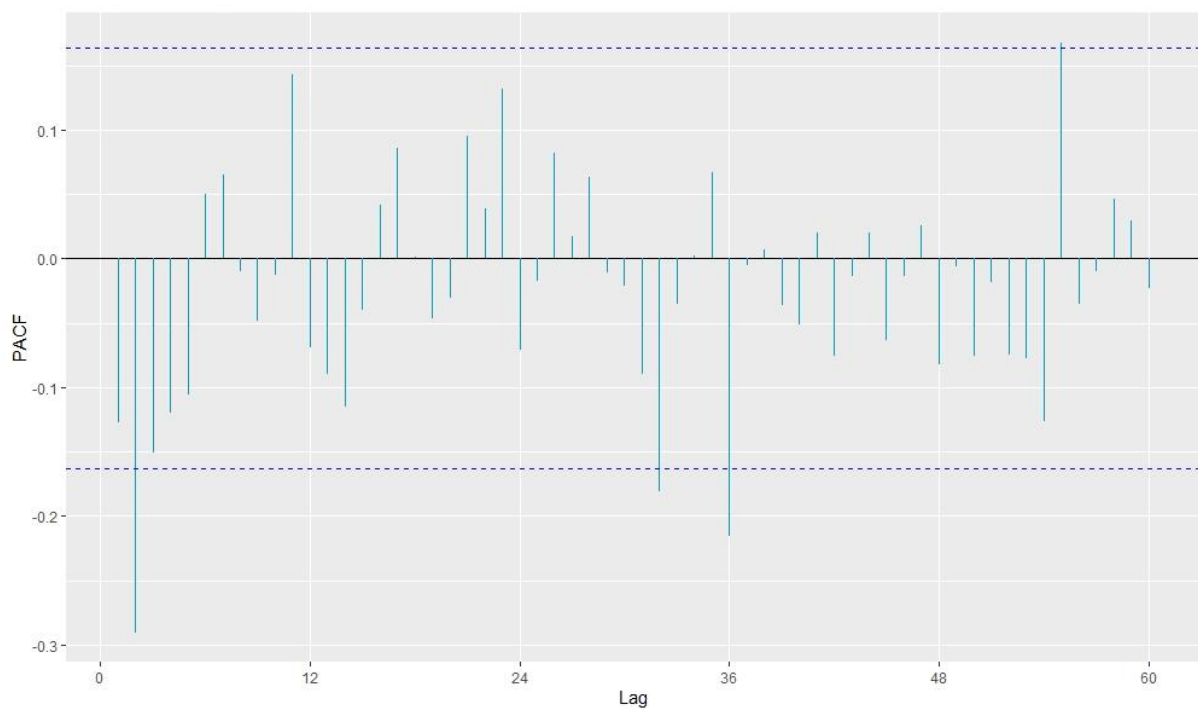


Figura 39. Función de autocorrelación parcial de los residuos del modelo  $SARIMA (1,1,0)(2,1,0)_{12}$

En las funciones de autocorrelación, tanto simple como parcial de los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$ , se observa que existen 3 retardos que sobresalen de los límites de confianza, pero aun así se deduce que los residuos cumplen con las condiciones de ruido blanco.

*Tabla 21.* Test de Ljung-Box de los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$

Chi-Cuadrado	GL	P-Valor
2.372	1	0.124

En la tabla 21 se observa el test de Ljung-Box, el cual arroja un valor chi-cuadrado de 2.372, por lo se determinó que no existen evidencias ( $p = 0.124$ ) para rechazar que todos los coeficientes de autocorrelación son simultáneamente igual a cero, lo que permite concluir que los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  se distribuyen de forma independiente.

*Tabla 22.* Test ANOVA de los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	12	12.2	0.021	0.885
Residuals	142	82545	581.3		

En la tabla 22 se observa el test ANOVA, el cual arroja un valor F de 0.021, por lo cual se puede determinar que no existen evidencias ( $p = 0.885$ ) para rechazar de que las perturbaciones son variables aleatorias de media nula, por lo que se concluye que los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  tienen medias igual a cero.

Tabla 23. Test de Levene de los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$

Df	Df	F value	Pr(>F)
11	132	2.593	0.005

En la tabla 23 se observa el test de Levene, el cual arroja un valor F de 2.593, en consecuencia se determinó que existen evidencias ( $p = 0.005$ ) para rechazar que todas las perturbaciones tienen las mismas varianzas, por lo que se concluye que los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  no tienen varianzas homogéneas.

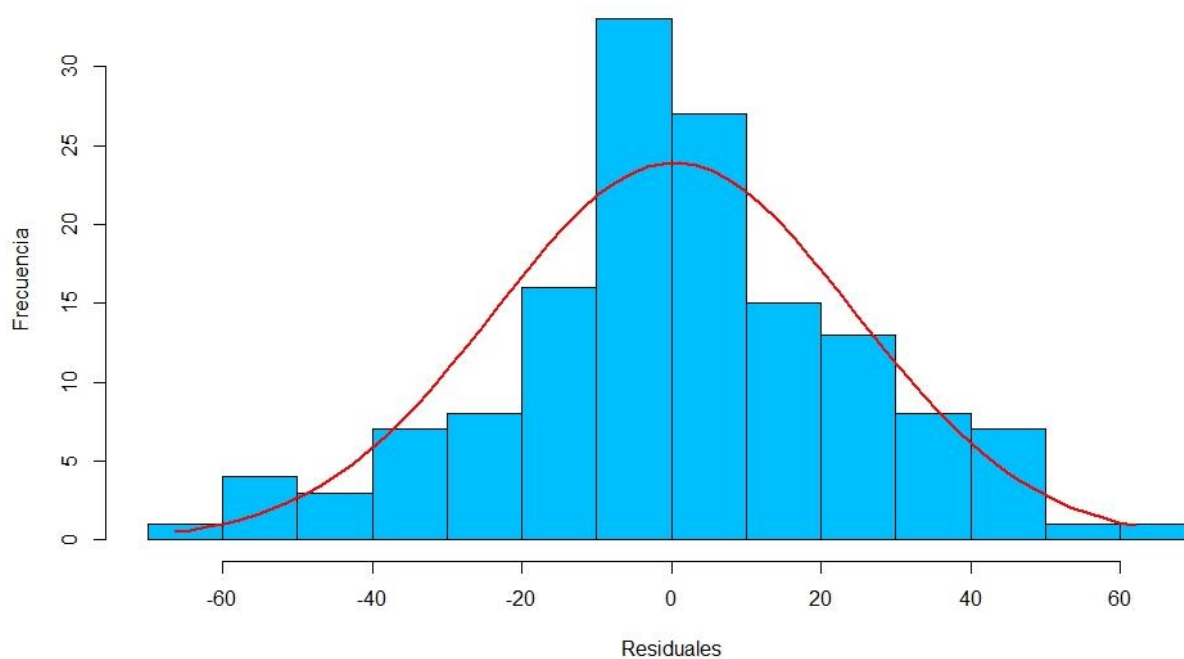


Figura 40. Histograma de los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$

En el histograma se observa que los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$ , tienden a seguir una distribución normal.

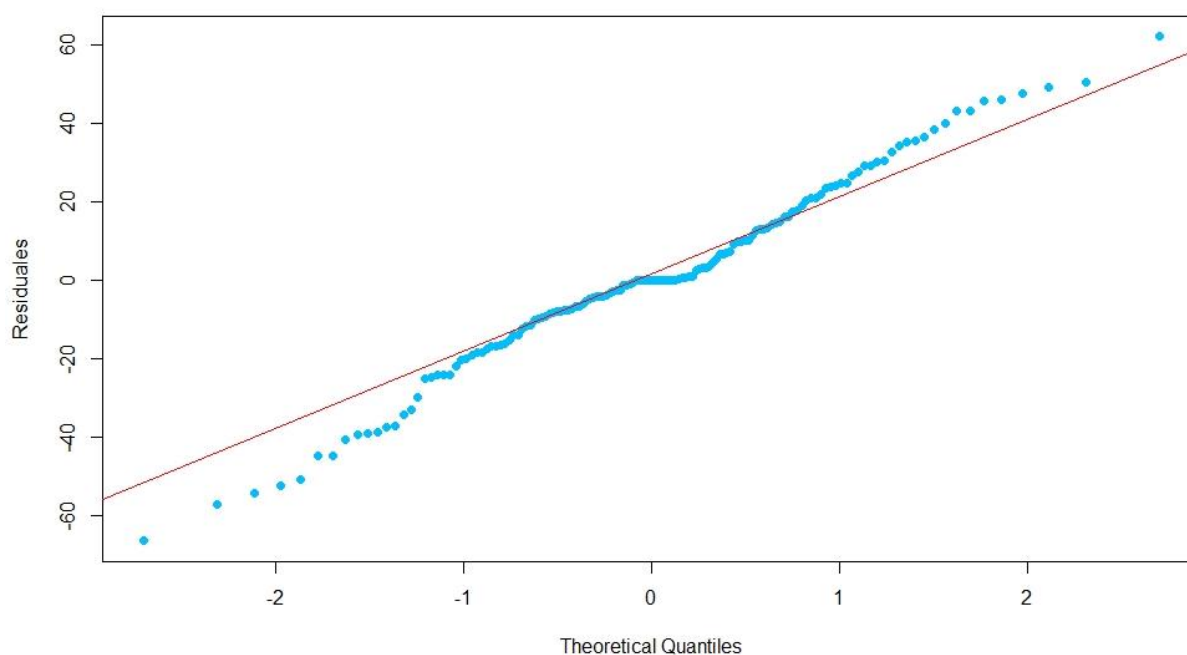


Figura 41. Diagrama Q-Q normal de los residuos del modelo  $SARIMA (1,1,0)(2,1,0)_{12}$

En el gráfico de probabilidad normal (diagrama Q-Q) muestra que los residuos del modelo  $SARIMA (1,1,0)(2,1,0)_{12}$ , siguen una línea recta, por lo tanto se deduce que siguen una distribución normal.

Tabla 24. Test de Kolmogorov-Smirnov de los residuos del modelo  $SARIMA (1,1,0)(2,1,0)_{12}$

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.080	0.321

En la tabla 24 se aprecia el estadístico de Kolmogorov - Smirnov, la cual arroja un valor de 0.080, por lo que se determinó que no existen evidencias ( $p = 0.321$ ) para rechazar la hipótesis de que las perturbaciones tienen una distribución conjunta normal, concluyendo que los residuos del modelo  $SARIMA (1,1,0)(2,1,0)_{12}$  siguen una distribución normal.

### MODELO SARIMA (2,1,0) (1,1,0)<sub>12</sub>

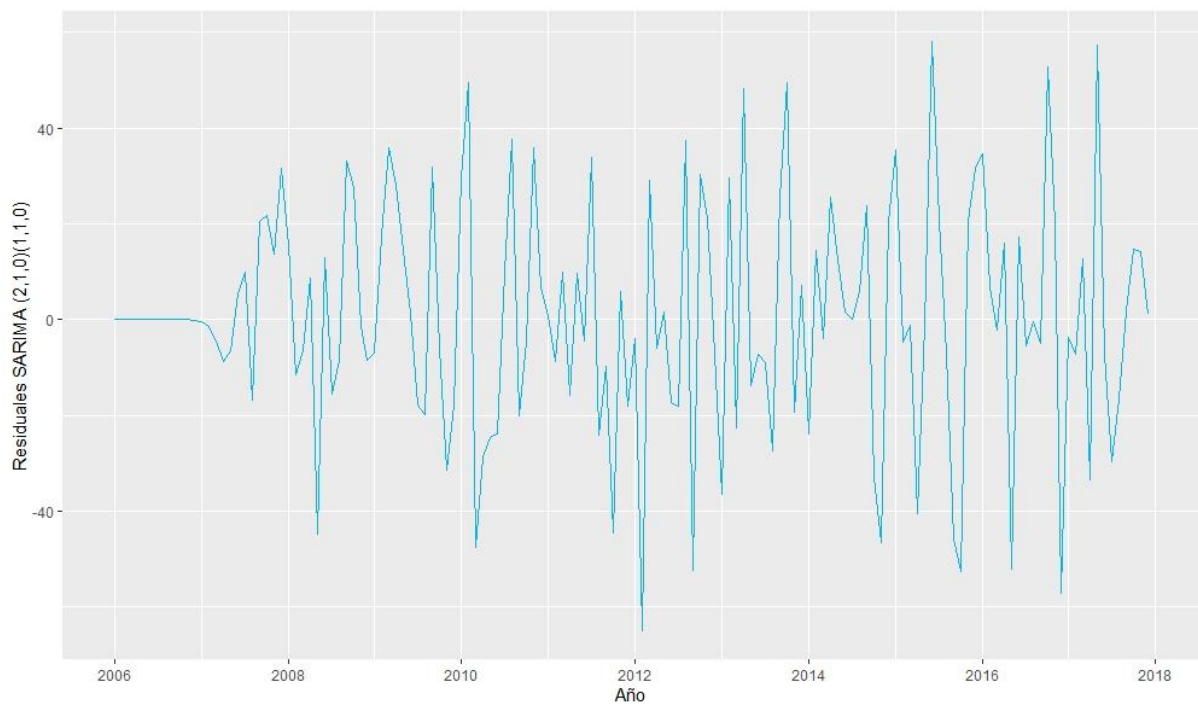


Figura 42. Residuos del modelo  $SARIMA (2,1,0)(1,1,0)_{12}$

En la figura 42 se presentan los residuos del modelo  $SARIMA (2,1,0)(1,1,0)_{12}$ , por la cual se deduce que siguen un ruido blanco.

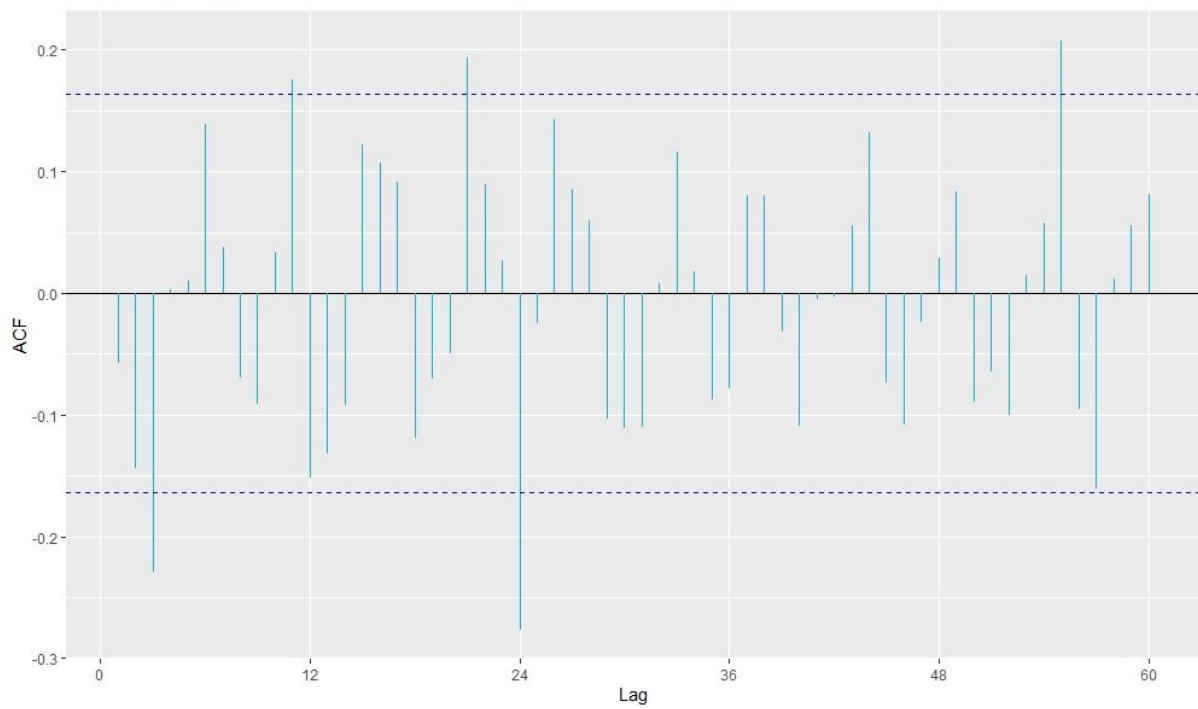


Figura 43. Función de autocorrelación de los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$

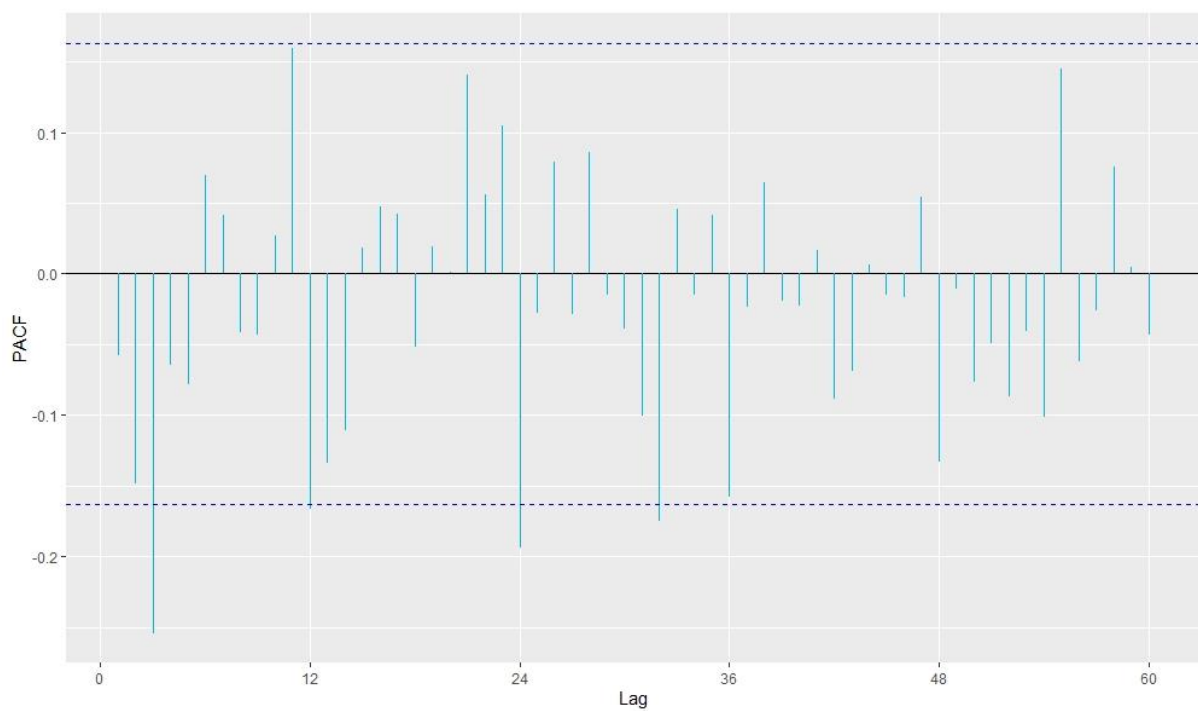


Figura 44. Función de autocorrelación parcial de los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$

En las funciones de autocorrelación, tanto simple como parcial, de los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$ , se observa que algunos retardos no están dentro de los límites de confianza, por lo que se deduce que los residuos no podrían cumplir con las condiciones de ruido blanco.

*Tabla 25.* Test de Ljung-Box de los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$

Chi-Cuadrado	GL	P-Valor
0.488	1	0.485

En la tabla 25 se observa el test de Ljung-Box, el cual arroja un valor chi-cuadrado de 0.488, por lo se determinó que no existen evidencias ( $p = 0.485$ ) para rechazar que todos los coeficientes de autocorrelación son simultáneamente igual a cero, lo que permite concluir que los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  se distribuyen de forma independiente.

*Tabla 26.* Test ANOVA de los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	0	0.1	0.000	0.989
Residuals	142	86003	605.7		

En la tabla 26 se observa el test ANOVA, el cual arroja un valor F de 0.000, por lo cual se puede determinar que no existen evidencias ( $p = 0.989$ ) para rechazar de que las perturbaciones son variables aleatorias de media nula, por lo que se concluye que los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  tienen medias igual a cero.

Tabla 27. Test de Levene de los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$

Df	Df	F value	Pr(>F)
11	132	0.799	0.641

En la tabla 27 se observa el test de Levene, el cual arroja un valor F de 0.799, en consecuencia se determinó que no existen evidencias ( $p = 0.641$ ) para rechazar que todas las perturbaciones tienen las mismas varianzas, por lo que se concluye que los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  tienen varianzas homogéneas.

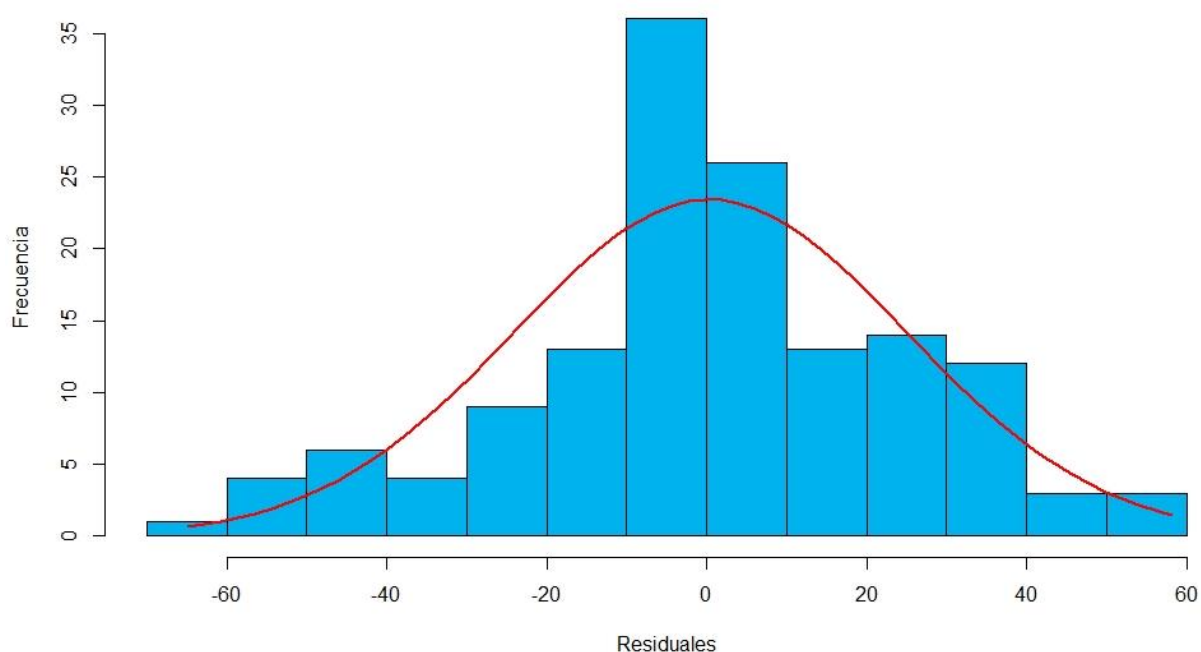


Figura 45. Histograma de los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$

En el histograma se observa que los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$ , tienden a seguir una distribución normal.



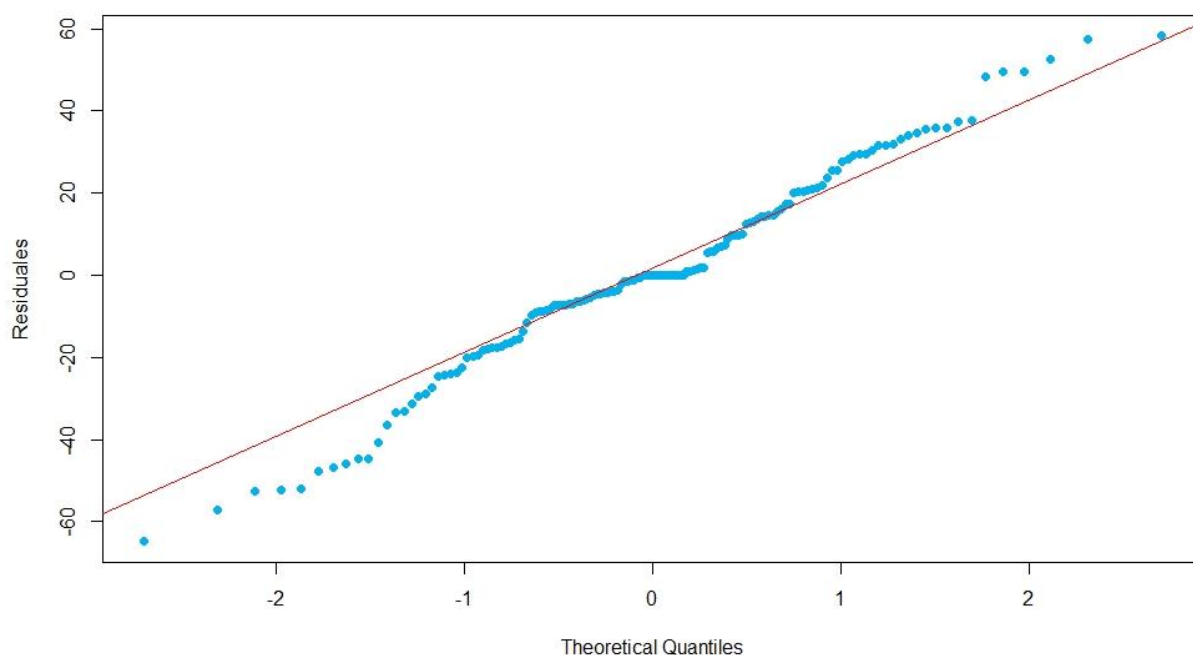


Figura 46. Diagrama Q-Q normal de los residuos del modelo  $SARIMA (2,1,0)(1,1,0)_{12}$

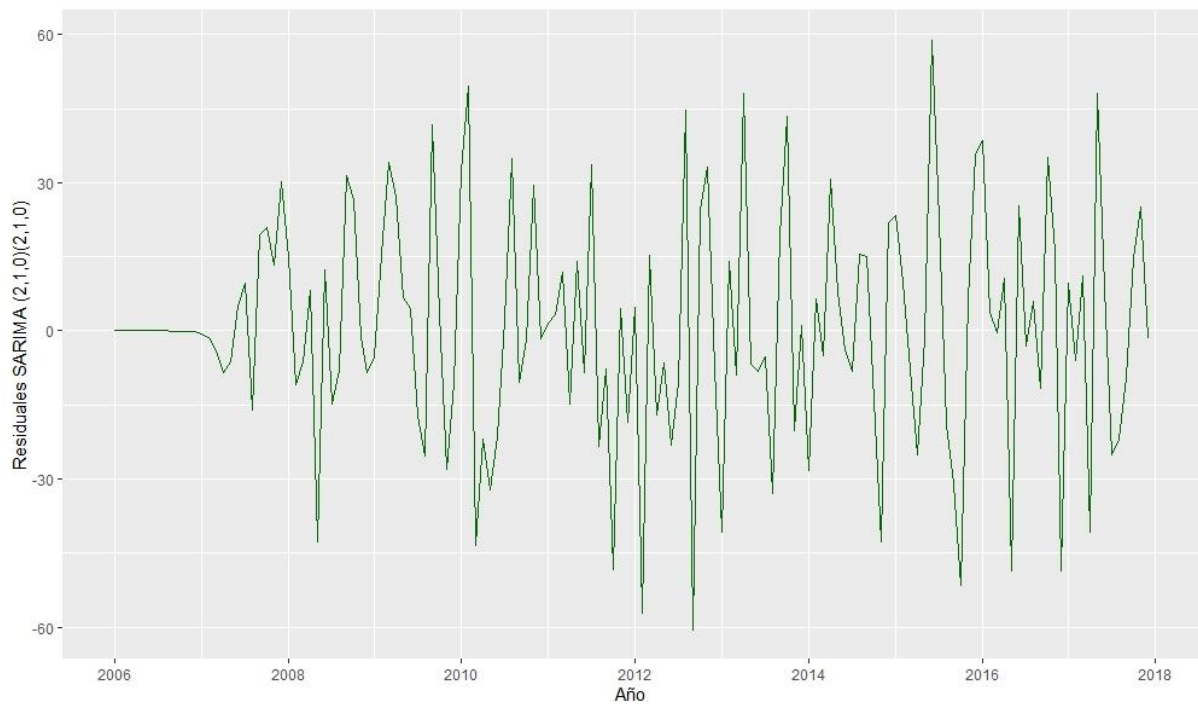
En el gráfico de probabilidad normal (diagrama Q-Q) muestra que los residuos del modelo  $SARIMA (2,1,0)(1,1,0)_{12}$ , siguen una línea recta, por lo tanto se deduce que siguen una distribución normal.

Tabla 28. Test de Kolmogorov-Smirnov de los residuos del modelo  $SARIMA (2,1,0)(1,1,0)_{12}$

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.080	0.321

En la tabla 28 se aprecia el estadístico de Kolmogorov - Smirnov, la cual arroja un valor de 0.080, por lo que se determinó que no existen evidencias ( $p = 0.321$ ) para rechazar la hipótesis de que las perturbaciones tienen una distribución conjunta normal, concluyendo que los residuos del modelo  $SARIMA (2,1,0)(1,1,0)_{12}$  siguen una distribución normal.

### MODELO SARIMA (2,1,0) (2,1,0)<sub>12</sub>



*Figura 47.* Residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$

En la figura 47 se tienen los residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$ , a través de la cual se deduce que los mismos siguen un ruido blanco.

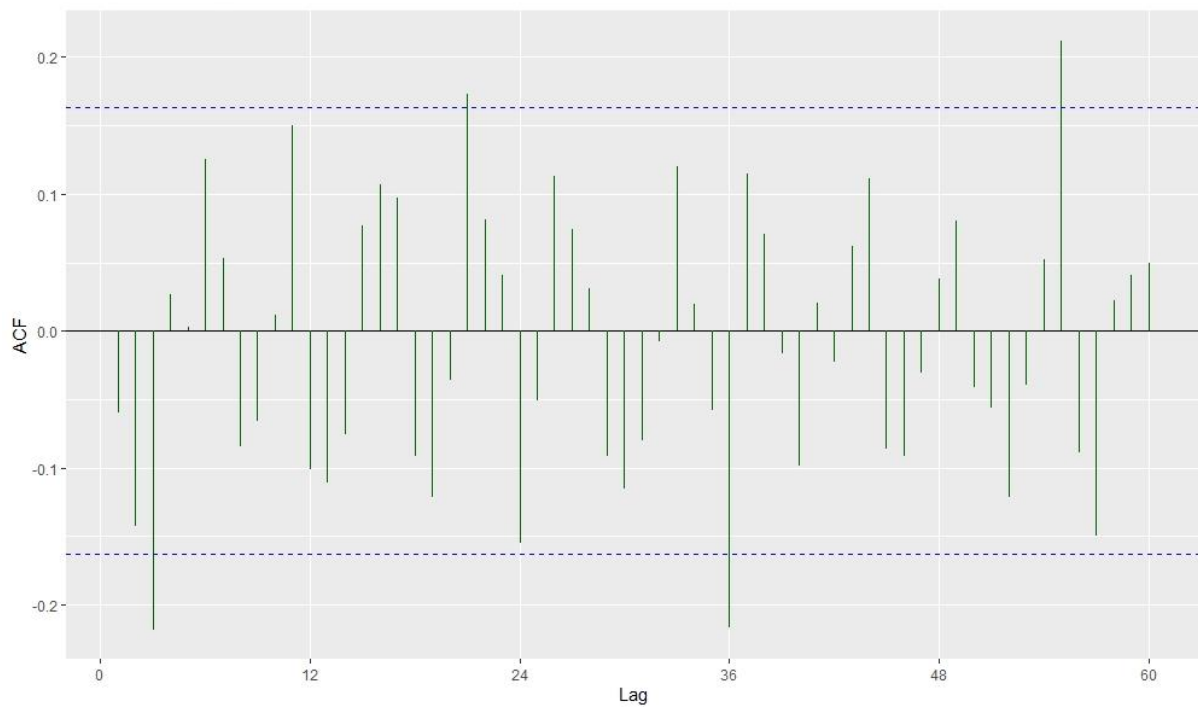


Figura 48. Función de autocorrelación de los residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$

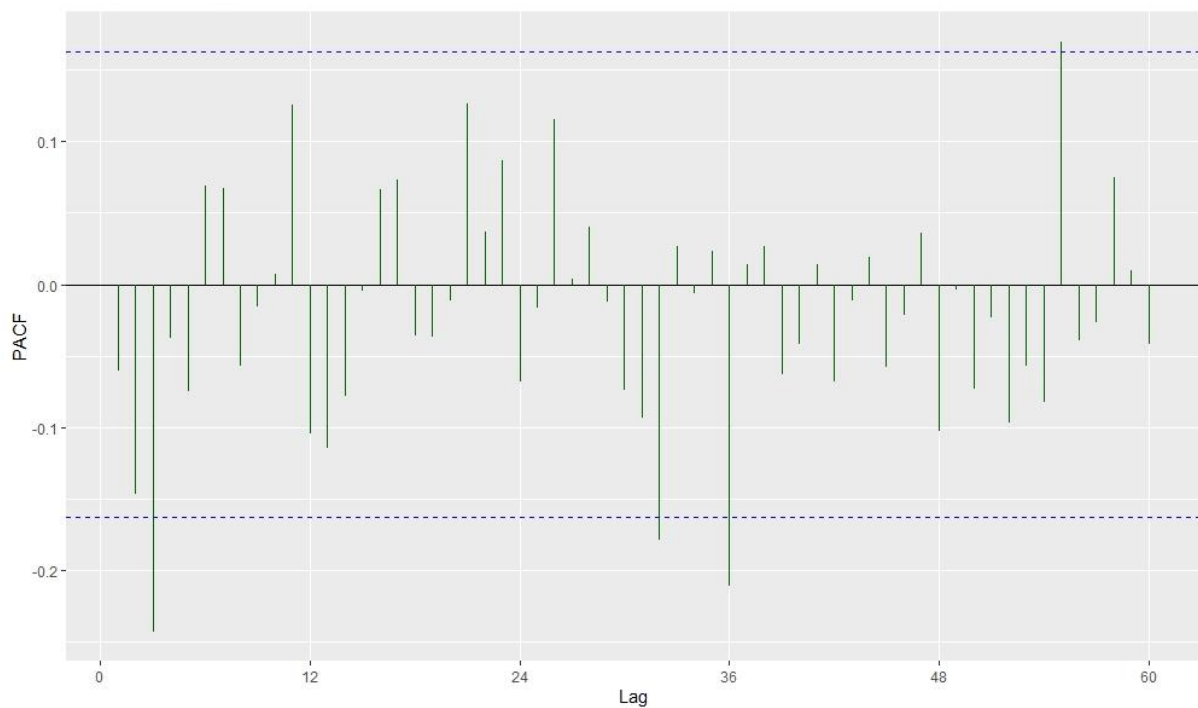


Figura 49. Función de autocorrelación parcial de los residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$

En las funciones de autocorrelación, tanto simple como parcial, de los residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$ , se observa que existen 2 retardos que no están dentro de los límites de confianza, pero aun así se deduce que los residuos cumplen con las condiciones de ruido blanco.

*Tabla 29.* Test de Ljung-Box de los residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$

Chi-Cuadrado	GL	P-Valor
0.531	1	0.466

En la tabla 29 se observa el test de Ljung-Box, el cual arroja un valor chi-cuadrado de 0.531, por lo se determinó que no existen evidencias ( $p = 0.466$ ) para rechazar que todos los coeficientes de autocorrelación son simultáneamente igual a cero, lo que permite concluir que los residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$  se distribuyen de forma independiente.

*Tabla 30.* Test ANOVA de los residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	20	20.3	0.037	0.848
Residuals	142	77899	548.6		

En la tabla 30 se observa el test ANOVA, el cual arroja un valor F de 0.037, por lo cual se puede determinar que no existen evidencias ( $p = 0.848$ ) para rechazar de que las perturbaciones son variables aleatorias de media nula, por lo que se concluye que los residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$  tienen medias igual a cero.

Tabla 31. Test de Levene de los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$

Df	Df	F value	Pr(>F)
11	132	2.419	0.009

En la tabla 31 se observa el test de Levene, el cual arroja un valor F de 2.419, en consecuencia se determinó que existen evidencias ( $p = 0.009$ ) para rechazar que todas las perturbaciones tienen las mismas varianzas, por lo que se concluye que los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  no tienen varianzas homogéneas.

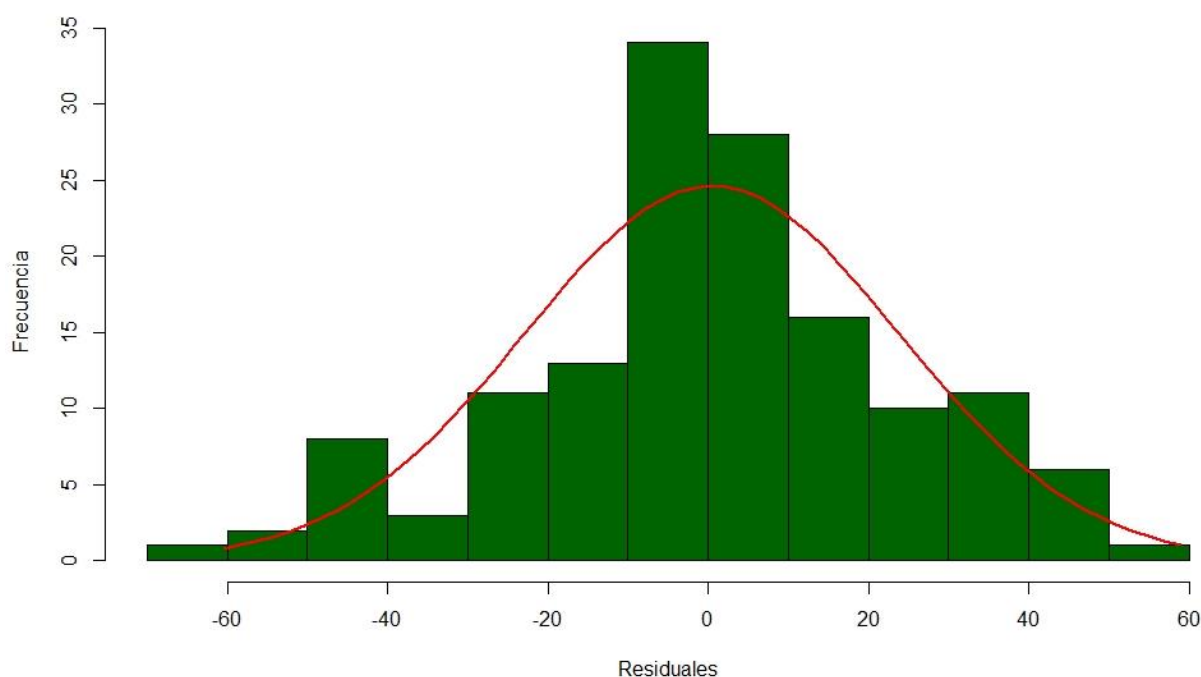


Figura 50. Histograma de los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$

En el histograma se observa que los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$ , tienden a seguir una distribución normal.

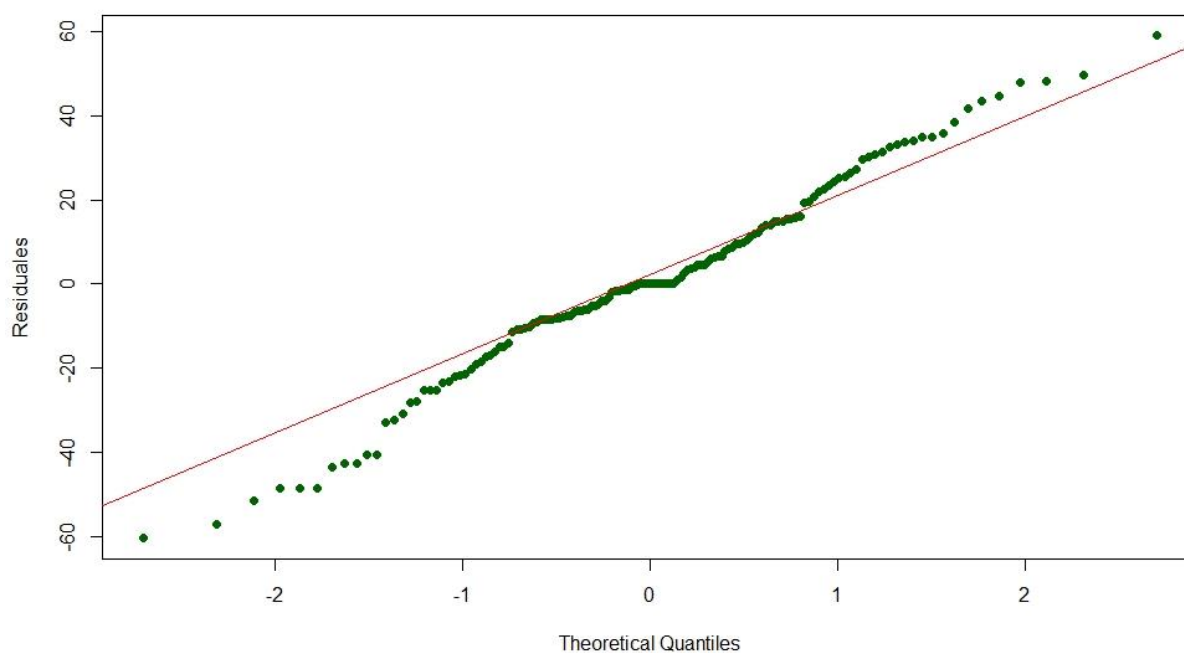


Figura 51. Diagrama Q-Q normal de los residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$

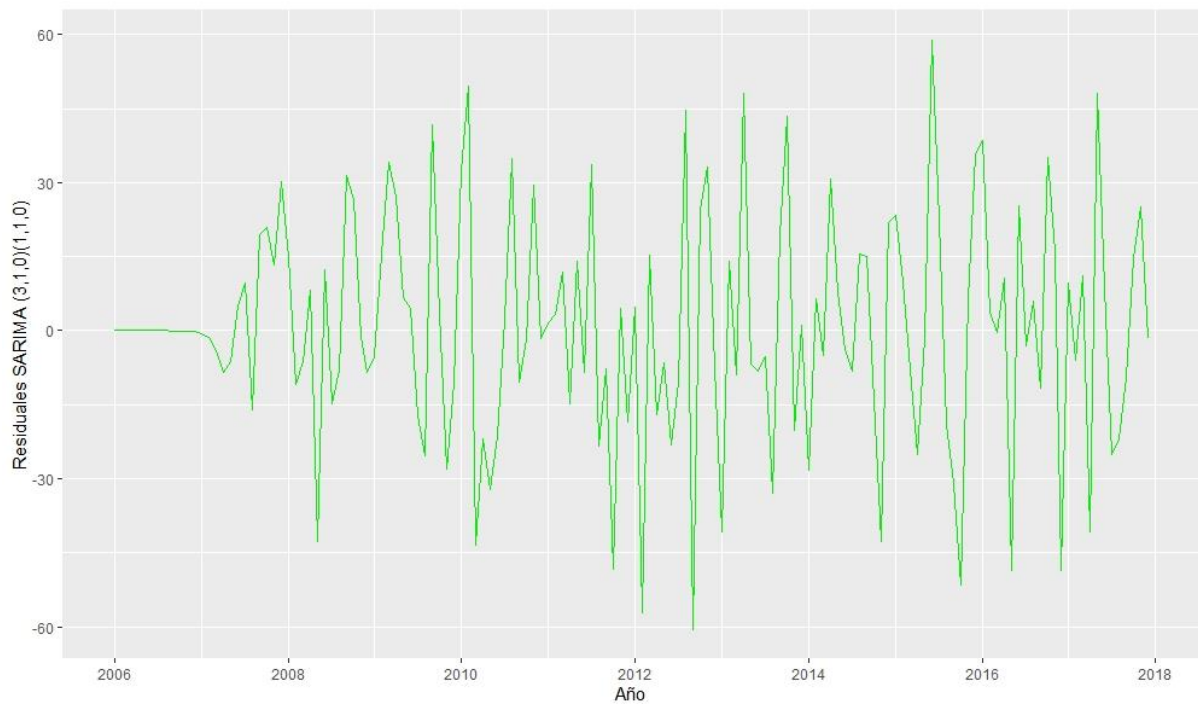
En el gráfico de probabilidad normal (diagrama Q-Q) muestra que los residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$ , siguen una línea recta, por lo tanto se deduce que siguen una distribución normal.

Tabla 32. Test de Kolmogorov-Smirnov de los residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.077	0.361

En la tabla 32 se aprecia el estadístico de Kolmogorov - Smirnov, la cual arroja un valor de 0.077, por lo que se determinó que no existen evidencias ( $p = 0.361$ ) para rechazar la hipótesis de que las perturbaciones tienen una distribución conjunta normal, concluyendo que los residuos del modelo  $SARIMA (2,1,0)(2,1,0)_{12}$  siguen una distribución normal.

### MODELO SARIMA (3,1,0) (1,1,0)<sub>12</sub>



*Figura 52.* Residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$

En la figura 52 se muestran los residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$ , a través de la cual se deduce que los residuos siguen un ruido blanco.

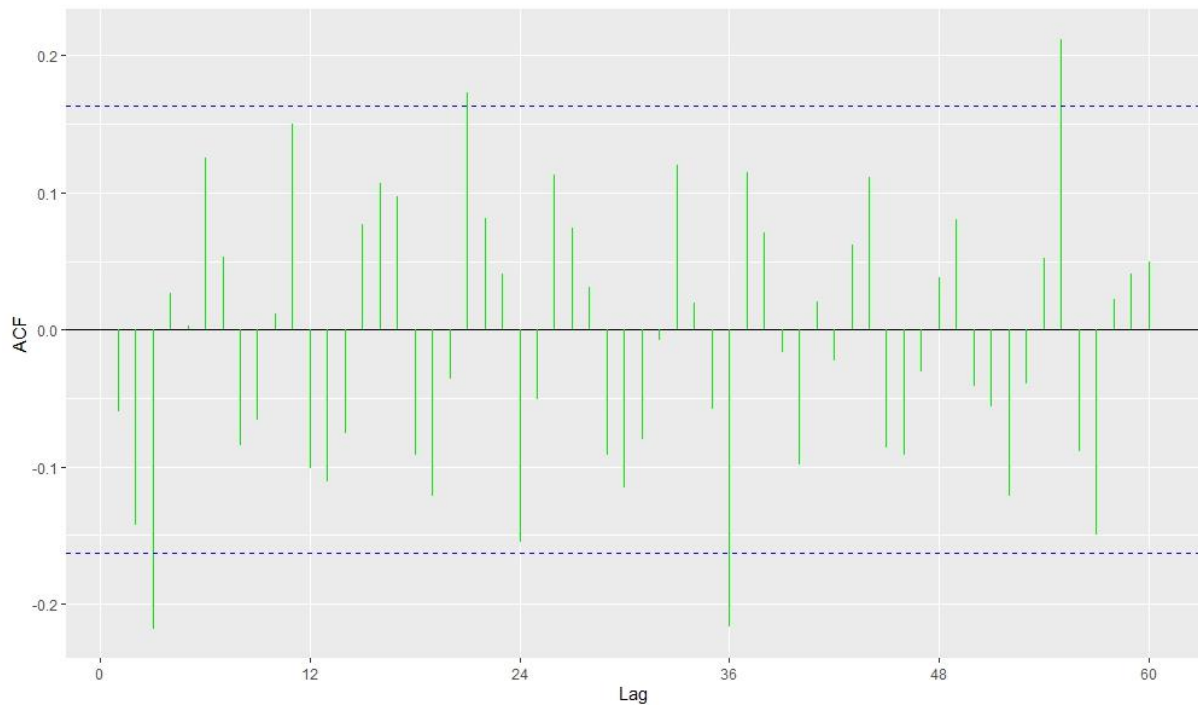


Figura 53. Función de autocorrelación de los residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$

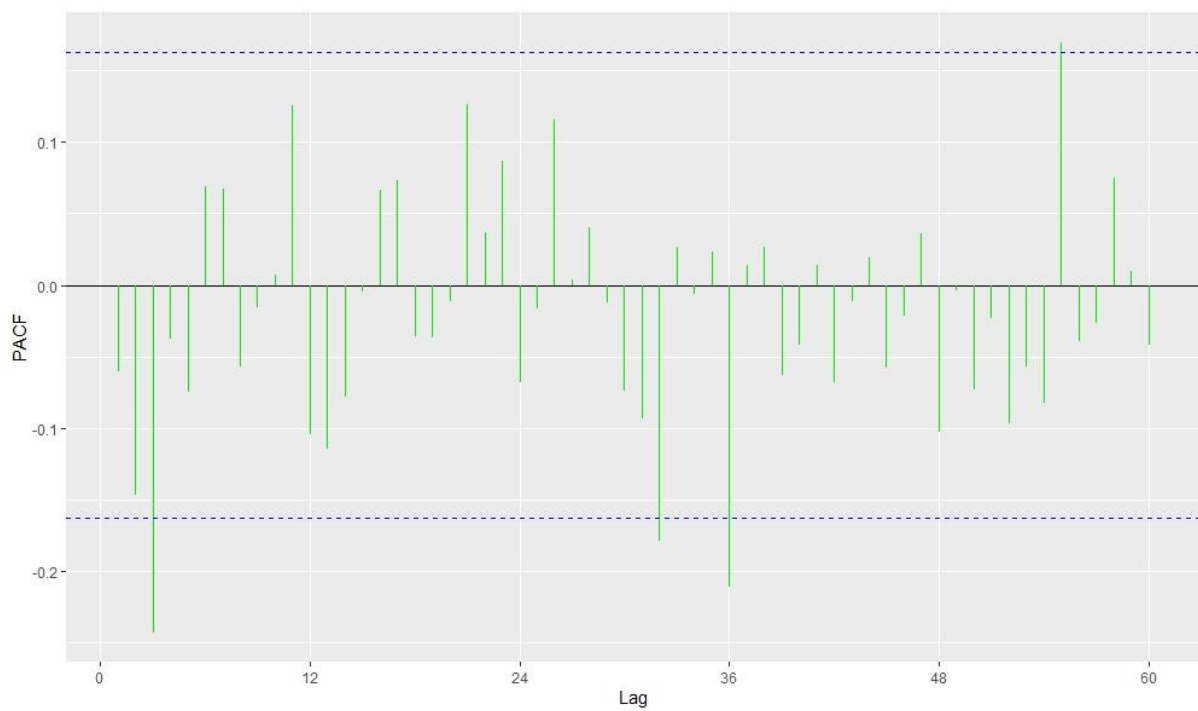


Figura 54. Función de autocorrelación parcial de los residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$



En las funciones de autocorrelación, tanto simple como parcial, de los residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$ , se observa que solo 3 retardos no están dentro de los límites de confianza, por lo que se deduce que los residuos cumplen con las condiciones de ruido blanco.

*Tabla 33.* Test de Ljung-Box de los residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$

Chi-Cuadrado	GL	P-Valor
0.234	1	0.629

En la tabla 33 se observa el test de Ljung-Box, el cual arroja un valor chi-cuadrado de 0.234, por lo se determinó que no existen evidencias ( $p = 0.629$ ) para rechazar que todos los coeficientes de autocorrelación son simultáneamente igual a cero, lo que permite concluir que los residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$  se distribuyen de forma independiente.

*Tabla 34.* Test ANOVA de los residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$

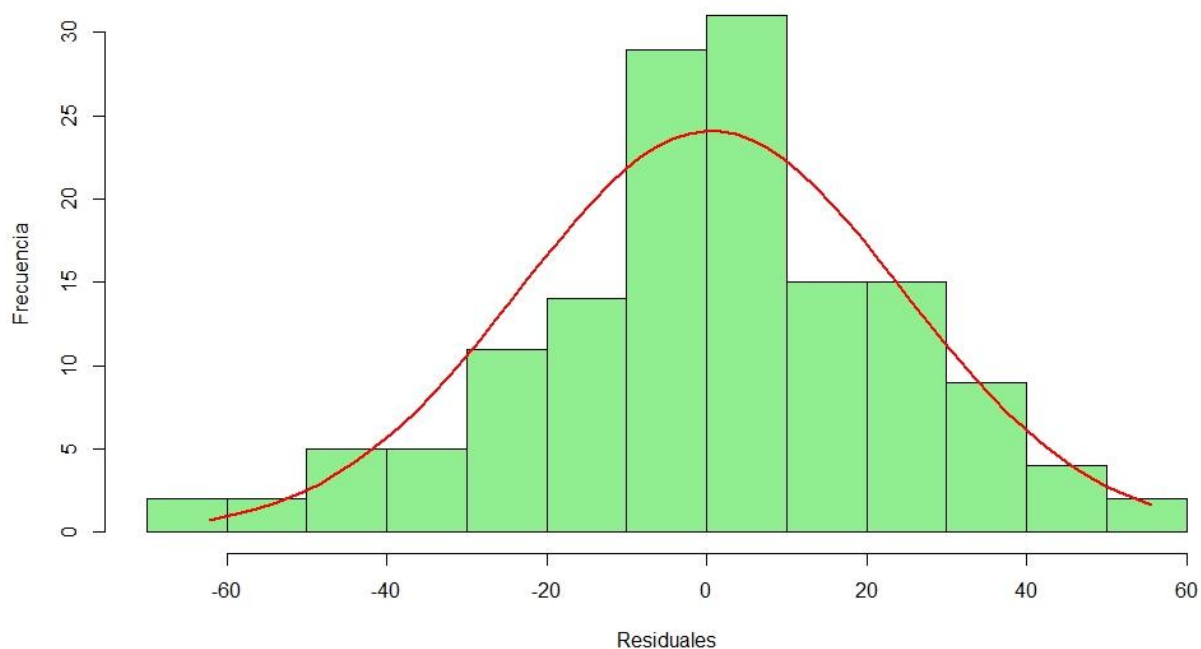
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$ año	1	34	34.1	0.060	0.807
Residuals	142	81172	571.6		

En la tabla 34 se observa el test ANOVA, el cual arroja un valor F de 0.060, por lo cual se puede determinar que no existen evidencias ( $p = 0.807$ ) para rechazar de que las perturbaciones son variables aleatorias de media nula, por lo que se concluye que los residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$  tienen medias igual a cero.

*Tabla 35. Test de Levene de los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$*

Df	Df	F value	Pr(>F)
11	132	2.518	0.007

En la tabla 35 se observa el test de Levene, el cual arroja un valor F de 2.518, en consecuencia se determinó que existen evidencias ( $p = 0.007$ ) para rechazar que todas las perturbaciones tienen las mismas varianzas, por lo que se concluye que los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  no tienen varianzas homogéneas.



*Figura 55. Histograma de los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$*

En el histograma se observa que los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$ , tienden a seguir una distribución normal.

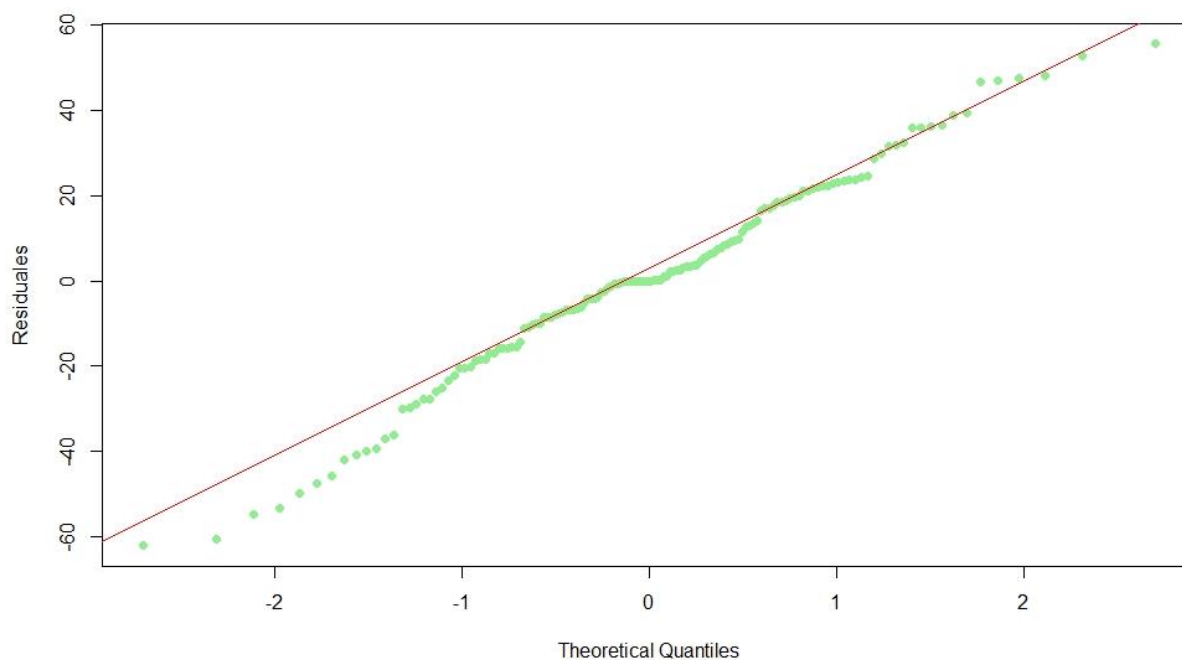


Figura 56. Diagrama Q-Q normal de los residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$

En el gráfico de probabilidad normal (diagrama Q-Q) muestra que los residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$ , siguen una línea recta, por lo tanto se deduce que siguen una distribución normal.

Tabla 36. Test de Kolmogorov-Smirnov de los residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.065	0.575

En la tabla 36 se aprecia el estadístico de Kolmogorov - Smirnov, la cual arroja un valor de 0.065, por lo que se determinó que no existen evidencias ( $p = 0.575$ ) para rechazar la hipótesis de que las perturbaciones tienen una distribución conjunta normal, concluyendo que los residuos del modelo  $SARIMA (3,1,0)(1,1,0)_{12}$  siguen una distribución normal.

### MODELO SARIMA (3,1,0) (2,1,0)<sub>12</sub>

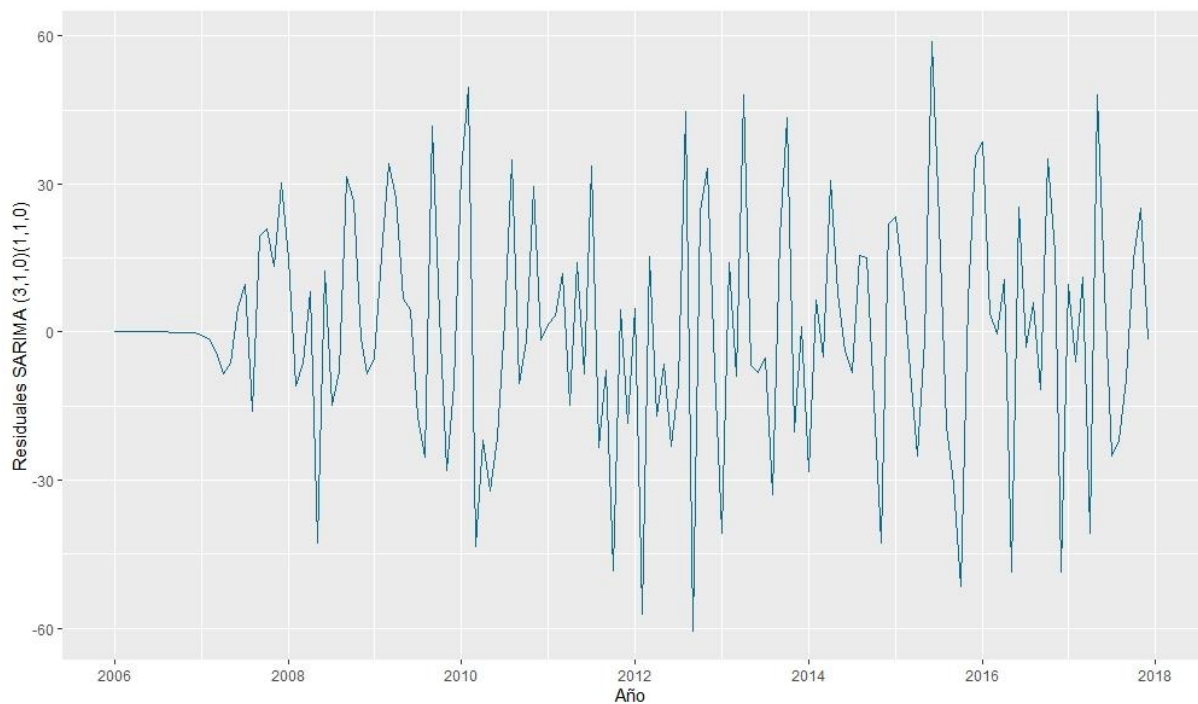


Figura 57. Residuos del modelo  $SARIMA (3,1,0)(2,1,0)_{12}$

En la figura 57 se muestran los residuos del modelo  $SARIMA (3,1,0)(2,1,0)_{12}$ , por medio de la cual se deduce que siguen un ruido blanco.

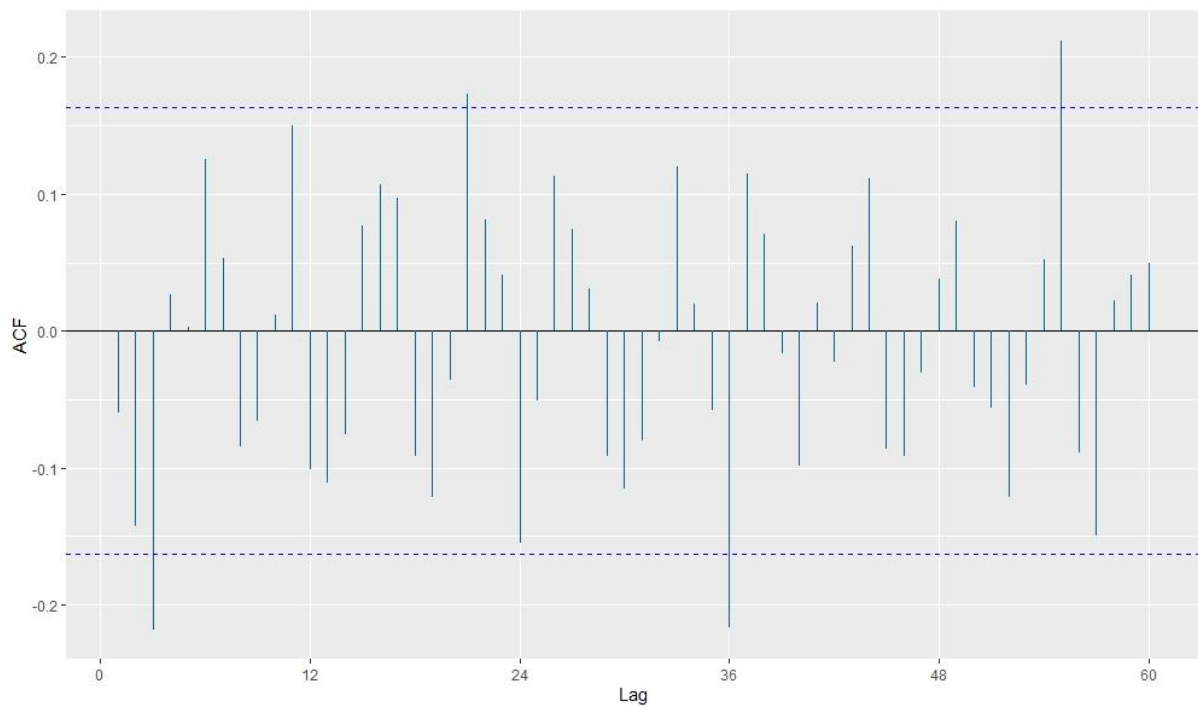


Figura 58. Función de autocorrelación de los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$

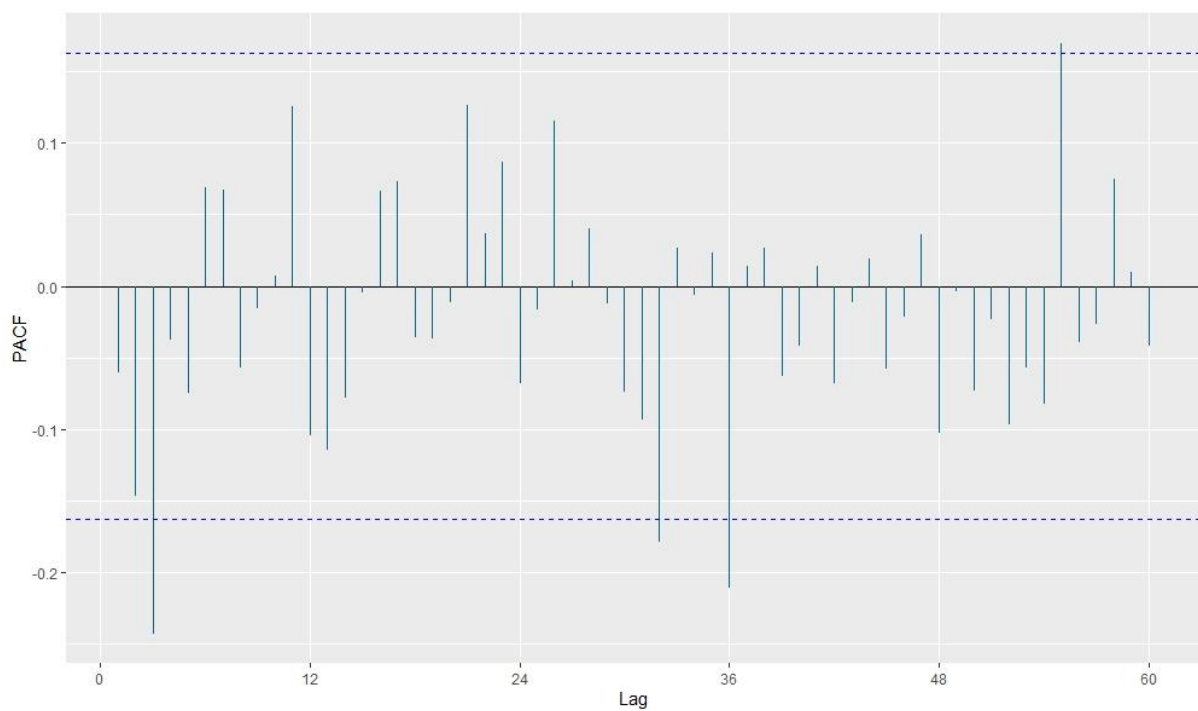


Figura 59. Función de autocorrelación parcial de los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$

En las funciones de autocorrelación, tanto simple como parcial, de los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$ , están dentro de los límites de confianza, por lo que se concluye que los residuos cumplen con las condiciones de ruido blanco.

*Tabla 37.* Test de Ljung-Box de los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$

Chi-Cuadrado	GL	P-Valor
0.205	1	0.651

En la tabla 37 se observa el test de Ljung-Box, el cual arroja un valor chi-cuadrado de 0.205, por lo se determinó que no existen evidencias ( $p = 0.651$ ) para rechazar que todos los coeficientes de autocorrelación son simultáneamente igual a cero, lo que permite concluir que los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  se distribuyen de forma independiente.

*Tabla 38.* Test ANOVA de los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	41	40.8	0.079	0.779
Residuals	142	73070	514.6		

En la tabla 38 se observa el test ANOVA, el cual arroja un valor F de 0.079, por lo cual se puede determinar que no existen evidencias ( $p = 0.779$ ) para rechazar de que las perturbaciones son variables aleatorias de media nula, por lo que se concluye que los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  tienen medias igual a cero.

Tabla 39. Test de Levene de los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$

Df	Df	F value	Pr(>F)
11	132	2.186	0.019

En la tabla 39 se observa el test de Levene, el cual arroja un valor F de 2.186, en consecuencia se determinó que existen evidencias ( $p = 0.019$ ) para rechazar que todas las perturbaciones tienen las mismas varianzas, por lo que se concluye que los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  no tienen varianzas homogéneas.

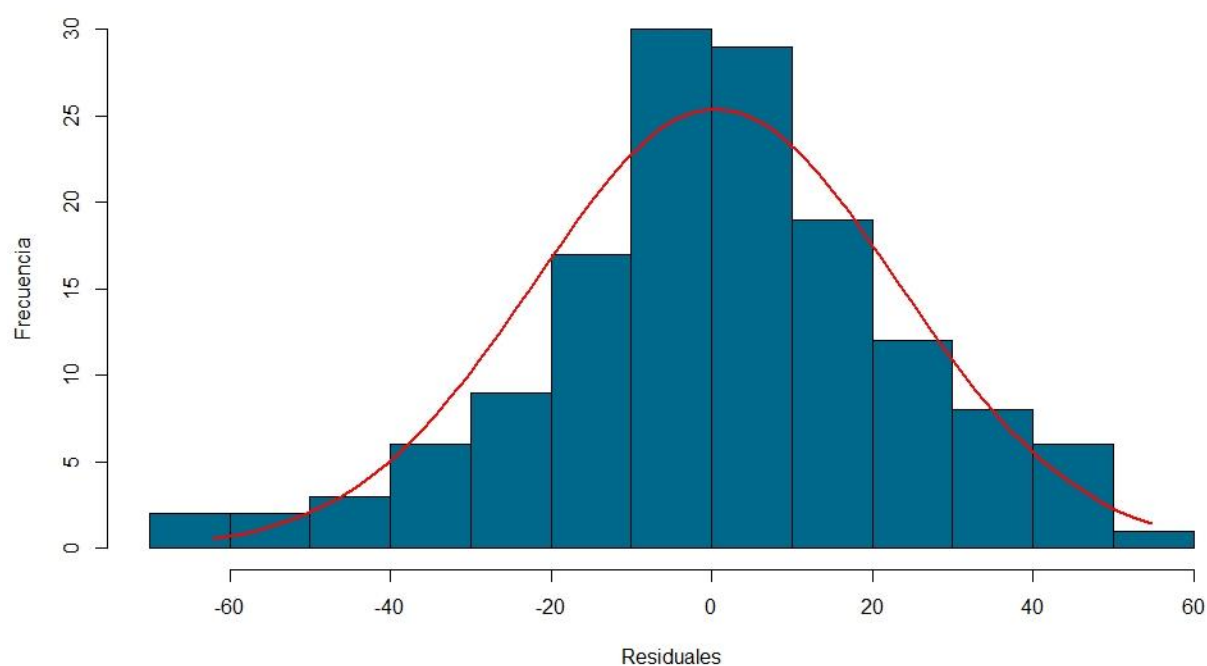


Figura 60. Histograma de los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$

En el histograma se observa que los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$ , tienden a seguir una distribución normal.

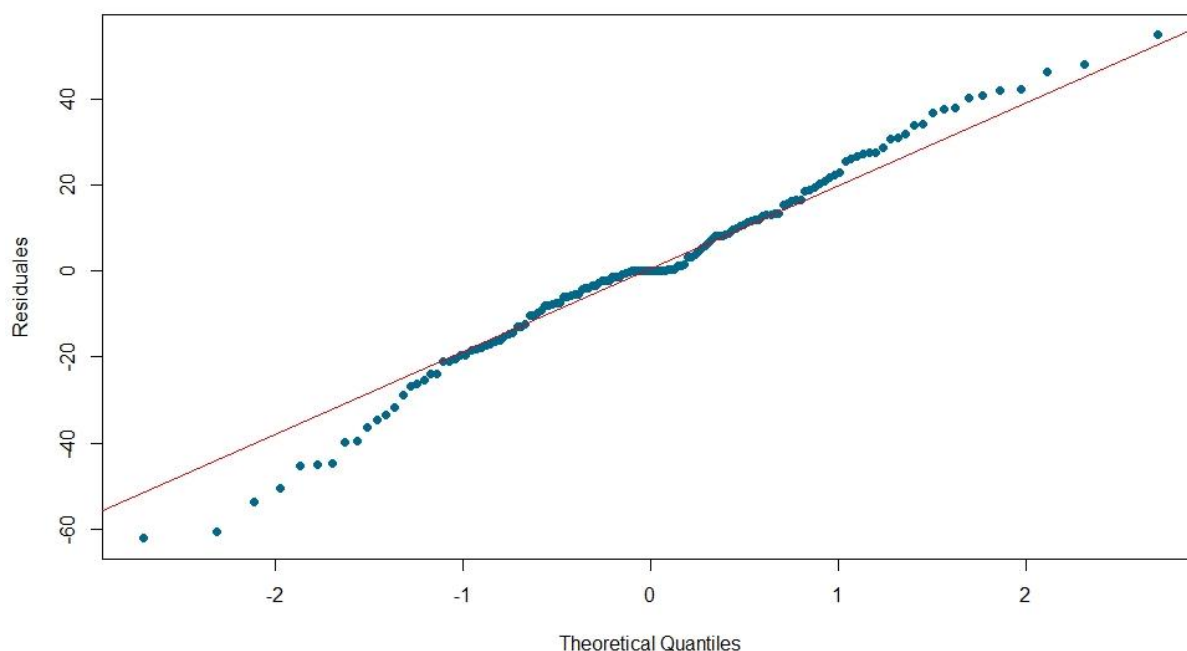


Figura 61. Diagrama Q-Q normal de los residuos del modelo  $SARIMA (3,1,0)(2,1,0)_{12}$

En el gráfico de probabilidad normal (diagrama Q-Q) muestra que los residuos del modelo  $SARIMA (3,1,0)(2,1,0)_{12}$ , siguen una línea recta, por lo tanto se deduce que siguen una distribución normal.

Tabla 40. Test de Kolmogorov-Smirnov de los residuos del modelo  $SARIMA (3,1,0)(2,1,0)_{12}$

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.066	0.554

En la tabla 40 se aprecia el estadístico de Kolmogorov - Smirnov, la cual arroja un valor de 0.066, por lo que se determinó que no existen evidencias ( $p = 0.554$ ) para rechazar la hipótesis de que las perturbaciones tienen una distribución conjunta normal, concluyendo que los residuos del modelo  $SARIMA (3,1,0)(2,1,0)_{12}$  siguen una distribución normal.



De los modelos propuestos, se concluye que modelos ARIMA (1,1,0), ARIMA (2,1,0), ARIMA (3,1,0) y SARIMA (2,1,0) (1,1,0) cumplen con los requisitos, los cuales son que sus residuos se distribuyan como un ruido blanco, sean independientes y sigan una distribución normal; ahora se realizara el cálculo de sus pronósticos para el año 2018.

*Tabla 41.* Pronósticos del número de pacientes con cáncer en el HNAAA con la metodología Box y Jenkins para el año 2018

	ARIMA (1,1,0)	ARIMA (2,1,0)	ARIMA (3,1,0)	SARIMA (2,1,0) (1,1,0)
Ene-18	173	176	172	201
Feb-18	163	170	171	189
Mar-18	169	169	172	193
Abr-18	166	171	168	195
May-18	167	170	171	190
Jun-18	166	170	170	214
Jul-18	167	170	170	186
Ago-18	166	170	170	189
Set-18	167	170	170	175
Oct-18	167	170	170	206
Nov-18	167	170	170	203
Dic-18	167	170	170	156

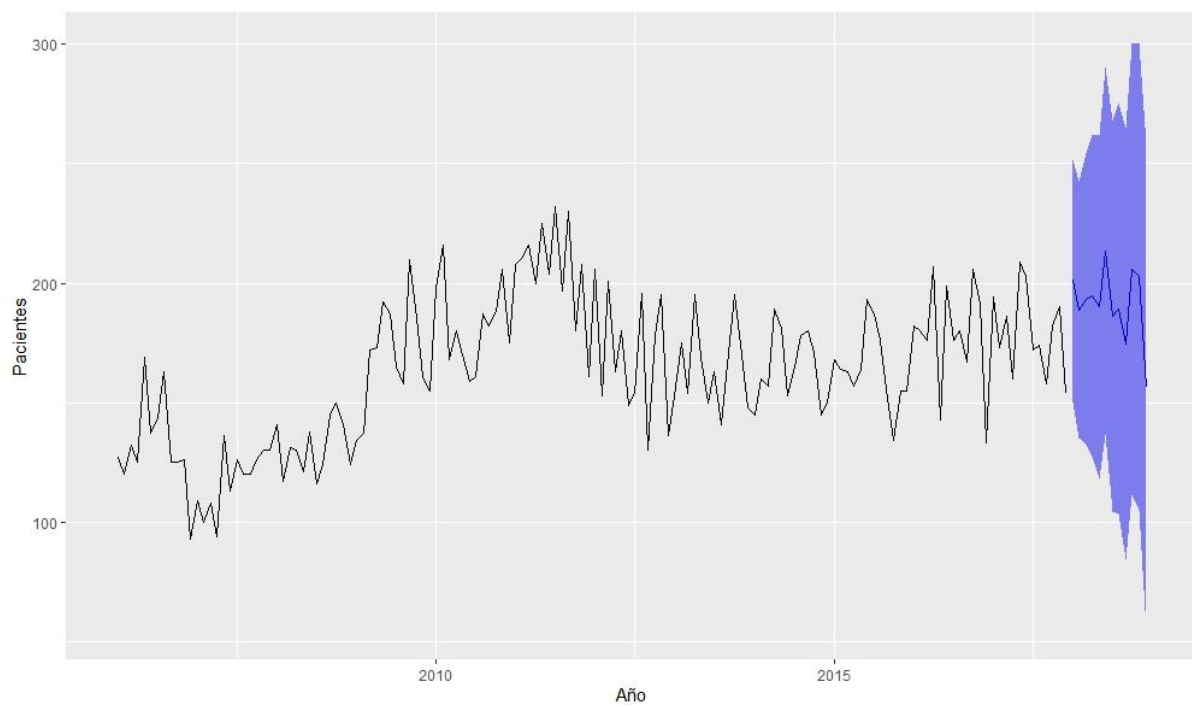
En la tabla 41 se aprecian los pronósticos del número de pacientes con cáncer en el HNAAA, según los modelos que describen la serie original.

*Tabla 42.* Comparación de la evaluación de los errores de pronósticos de los modelos propuestos

	RMSE	MAE	MAPE	AIC	BIC
ARIMA (1,1,0)	21.855	17.408	10.686	1293.300	1299.220
ARIMA (2,1,0)	21.144	17.298	10.620	1285.980	1294.870
ARIMA (3,1,0)	20.673	16.533	10.223	1281.650	1293.500
SARIMA (2,1,0) (1,1,0)	24.443	18.439	11.268	1233.090	1244.590

En la tabla 42 se tiene la evaluación de los errores de pronósticos de todos los modelos propuestos, en la cual se ha resaltado los menores valores con respecto al error mencionado. A través de este análisis de los errores se observa que el modelo *SARIMA* (2,1,0)(1,1,0)<sub>12</sub> es el modelo que tiene buenos criterios de información como son los de Akaike y Bayesiano, además de tener menores errores.

SARIMA (2,1,0) (1,1,0)<sub>12</sub>



*Figura 62.* Pronósticos con el modelo SARIMA (2,1,0)(1,1,0)<sub>12</sub> del número de pacientes con cáncer en el HNAAA para el año 2018

### 3.3 Identificación del modelo de serie de tiempo para el número de pacientes con cáncer, mediante la metodología de redes neuronales.

*Tabla 43.* Pesos sinápticos de la primera capa oculta de la red neuronal

Pesos sinápticos																	
-0.429	1.439	0.070	0.356	-0.774	1.148	0.506	-0.451	0.660	0.585	-1.155	-0.949	-0.706	0.148	-1.007	0.558	0.536	1.178
-0.025	-0.521	0.428	-2.909	-1.219	-0.491	0.884	3.431	-3.611	-1.265	-0.992	0.045	-2.460	1.297	1.845	-0.052	2.016	-0.243

En la presente tabla se observan los 36 pesos sinápticos de la primera capa oculta de la red neuronal, la cual se puede observar que hay 18 pesos positivos, los cuales actúan como activadores de las neuronas, mientras que 18 pesos son negativos y actúan como inhibidores.

*Tabla 44.* Pesos sinápticos de la segunda capa oculta de la red neuronal

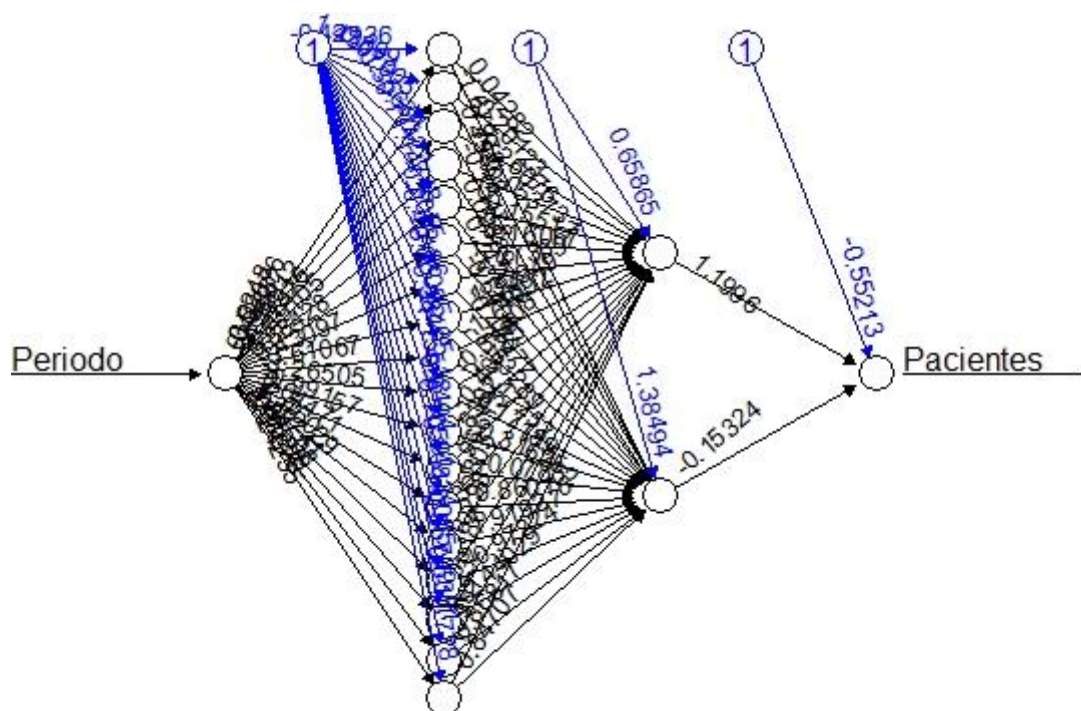
Pesos sinápticos																		
0.659	0.043	-0.201	-0.261	-1.054	-0.155	-1.161	2.274	1.340	-1.741	-0.344	-0.408	0.839	-0.966	0.358	0.103	0.062	1.488	-0.109
1.385	-1.680	0.453	-0.155	0.099	0.248	-0.641	-0.471	-1.633	-1.178	0.172	0.316	-0.079	0.860	1.919	-0.972	-2.440	0.868	-0.847

En esta tabla se observan los 38 pesos sinápticos de la segunda capa oculta de la red neuronal, la cual se puede observar que hay 18 pesos positivos, los cuales actúan como activadores de las neuronas, mientras que 20 pesos son negativos y actúan como inhibidores de la red.

*Tabla 45.* Pesos sinápticos de la tercera capa oculta de la red neuronal

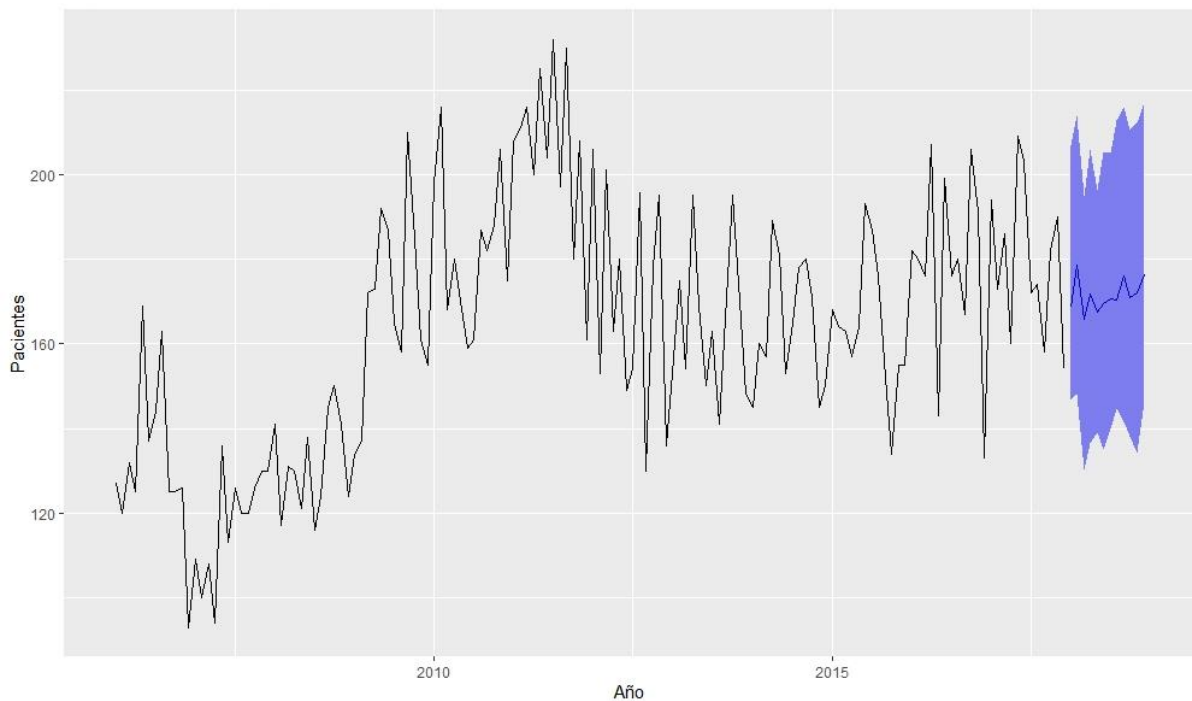
Pesos sinápticos
-0.5521
1.1996
-0.1532

En la tabla 45 se observan los 3 pesos sinápticos de la tercera capa oculta de la red neuronal, la cual se puede observar que hay 1 peso positivo, el cual actúa como activador de la neurona, mientras que 2 pesos son negativos y actúan como inhibidores.



*Figura 63.* Arquitectura de Redes Neuronales Artificiales (RNA) del número de pacientes con cáncer en el HNAAA, Chiclayo 2006 – 2017

En la figura 63 se observa la arquitectura RNA, la cual cuenta con 2 capas ocultas: 1 capa con 18 neuronas y otra capa con 2 neuronas. Dicha red sigue una propagación hacia adelante, lo que quiere decir es que el nivel de las neuronas, están conectadas a las neuronas del siguiente nivel.



*Figura 64.* Pronóstico del número de pacientes con cáncer en el HNAAA, mediante el algoritmo de redes neuronales, para el año 2018

En la figura 64 se observa los pronósticos calculados para el año 2018 utilizando redes neuronales artificiales, estos pronósticos siguen la tendencia de los años anteriores, lo cual evidencia que el procedimiento seguido y los valores estimados obtenidos son adecuados.

Así mismo dichos pronósticos sirven para realizar la comparación del mejor modelo mediante la metodología de Box y Jenkins, que a continuación se presentan:

### 3.4 Pronóstico del número de pacientes con cáncer para el año 2018, mediante la metodología de Box y Jenkins y redes neuronales

*Tabla 46.* Pronósticos del número de pacientes con cáncer en el HNAAA para el año 2018, usando la metodología de Box y Jenkins y redes neuronales

Fecha	SARIMA (2,1,0) (1,1,0)	Límite inferior 95%	Límite superior 95%	Redes Neuronales	Límite inferior 95%	Límite superior 95%
Ene-18	201	150	252	169	137	206
Feb-18	189	135	243	179	146	207
Mar-18	193	133	254	166	137	203
Abr-18	195	128	262	172	135	205
May-18	190	118	262	167	136	211
Jun-18	214	137	291	169	144	199
Jul-18	186	105	268	171	143	215
Ago-18	189	103	275	170	138	210
Sep-18	175	84	265	176	144	215
Oct-18	206	112	300	171	140	213
Nov-18	203	105	301	172	143	222
Dic-18	156	55	258	176	144	217

En la tabla 46 se aprecia los valores pronosticados para el número de pacientes con cáncer en el HNAAA., para el año 2018; también se presentan los límites de confianza al 95% para cada mes pronosticado.

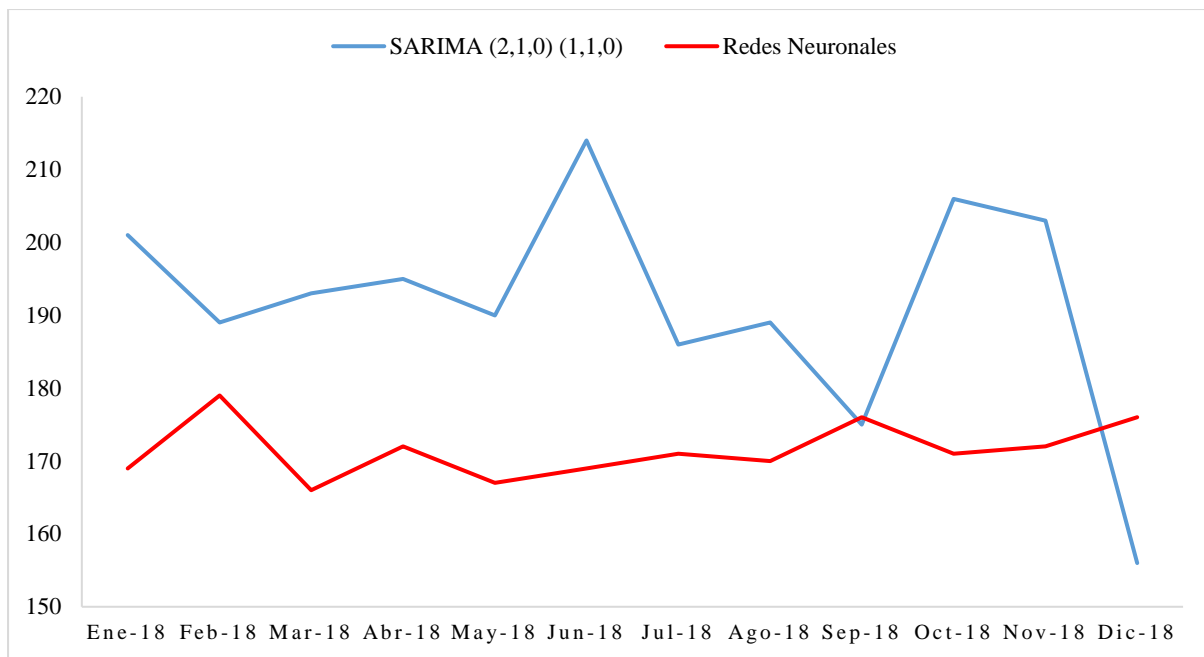
### 3.5 Comparación del mejor pronóstico que se ajusta al número de pacientes con cáncer

*Tabla 47.* Comparación de los modelos de redes neuronales y SARIMA

	RMSE	MAE	MAPE	MASE
SARIMA (2,1,0) (1,1,0)	24	18	11.27	0.75
RNA	14	11	6.62	0.44

En la presente tabla se muestra indicadores para identificar el mejor modelo, tales como la raíz del error cuadrático medio (RMSE). Medida de cuanto se desvía la serie dependiente del nivel pronosticado por el modelo, para el modelo de redes neuronales es 14 pacientes y para el modelo SARIMA es 24. El MAPE, error absoluto porcentual promedio, también mide la desviación de la serie dependiente del nivel pronosticado por el modelo, en este caso para redes neuronales es 6.62% y para el modelo SARIMA es 11.27%, siendo este último mucho mayor que el modelo de redes neuronales.





*Figura 65.* Pronóstico del número de pacientes con cáncer en el HNAAA, mediante el modelo SARIMA y RNA, para el año 2018

En esta figura se aprecia la comparación de los pronósticos obtenidos con el modelo SARIMA y con RNA, se observan que ambos valores pronosticados se ajustan a la serie en estudio, pero los pronósticos de redes neuronales son los que mejor se ajustan.

## CAPÍTULO IV. CONCLUSIONES

1. La serie original presenta tendencia y no es estacionaria.
2. El modelo estimado mediante la metodología de Box y Jenkins, para el número de pacientes con cáncer en el Hospital Nacional Almanzor Aguinaga Asenjo. Periodo 2006 – 2017, fue el modelo  $SARIMA (2, 1, 0) (1, 1, 0)_{12}$ .
3. Mediante la aplicación de redes neuronales artificiales (RNA), para el pronóstico del número de pacientes con cáncer en el Hospital Nacional Almanzor Aguinaga Asenjo. Periodo 2006 – 2017; el modelo es  $NNAR (4, 1, 3)_{12}$ ; donde 4 son las covariables, es decir todos los trimestres del año; 1 capa oculta, 3 nodos en la capa oculta.
4. Con la metodología de Box y Jenkins se estima que para el año 2018, habrá 2297 pacientes con cáncer en el HNAAA; mientras que con RNA habrá 2058 pacientes.
5. El modelo de redes neuronales es el modelo que mejor se ajusta a la serie número de pacientes con cáncer en el Hospital Nacional Almanzor Aguinaga Asenjo, por tener menor error cuadrático medio.

## REFERENCIAS BIBLIOGRÁFICAS

1. Alam, M., Kumar, M., Chowdhury, M., Siddique, A. (2015). An Algorithm for Training Multilayer Perceptron (MLP) For Image Reconstruction Using Neural Network without Overfitting. *International Journal of Scientific and Technology Research*, 4(2), 271-275.
2. Alsmadi, M., Omar, K., y Noah, S. (2009). Backpropagation algorithm: The best algorithm among the multi-layer perceptron algorithm. *International Journal of Computer Science and Network Security*, 9(4), 378–383.
3. Barrio, D. (5 de julio de 2017). El cáncer de mama sigue siendo una preocupación en el Perú. El Comercio. Recuperado de <http://elcomercio.pe>
4. Barrio, D. (11 de julio de 2017). Cáncer: el alto riesgo de morir si naciste en el Perú. El Comercio. Recuperado de <http://elcomercio.pe>
5. Cardenas, J. (2015). Pronósticos y comparación de una serie de tiempo con cambios estructurales mediante la red neuronal artificial de retropropagación resiliente y modelos no lineales (tesis de pregrado). Universidad Nacional Mayor de San Marcos, Lima, Perú.
6. Carlos, S., Lima, A., Brito, C., Rodrigues, J., Escórcio, L., Monte, L. y Teixeira, M. (2012). *Oncologia Básica*. Teresina, Brasil: Fundação Quixote
7. Castillo, R., & Varela, R. (2011). *Econometría práctica: Fundamentos de serie de tiempo*. Universidad Autónoma de Baja California, Baja California, México.
8. Chatfield, C., & Xing, H. (2019). *The Analysis of Time Series: An Introduction with R*. Florida, EE.UU: Chapman & Hall/CRC.
9. Correo (11 de Octubre del 2017). Más de 400 casos de cáncer en lo que va del año en Lambayeque. Correo. Recuperado de <https://diariocorreio.pe>

10. Cowpertwait, P. y Metcalfe, A. (2009). *Introductory Time Series with R*, EE.UU: Springer.
11. Cruz, I. (2007). *Pronósticos en el mercado de derivados utilizando redes neuronales y modelos ARIMA: una aplicación al Cete de 91 días en el MexDer* (tesis de maestría). Universidad Nacional Autónoma de México, Coyoacán, México.
12. Damian, M., Sandoval, N. (2018). *Modelo óptimo de pronóstico del índice mensual de producción de electricidad* (tesis de pregrado). Universidad Nacional Pedro Ruiz Gallo, Lambayeque, Perú.
13. Enders, W. (2015). *Applied econometric time series*. New Jersey, EE.UU: Wiley.
14. Gestión (5 de febrero de 2016). *Cáncer en Perú: El 85% de casos se detectan en estadios avanzados*. Gestión. Recuperado de <http://gestion.pe>
15. Gómez, W. (2010). *Pronóstico de una serie temporal usando redes neuronales* (tesis de pregrado). Universidad Tecnológica de la Mixteca, Oaxaca, México.
16. Goss, P., Lee, B., Badovinac, T., Strasser, K., Chavarri, Y., St Louis, J.,..., Azenha, G. (2013). *La planificación del control del cáncer en América Latina y el Caribe*, *Lancet Oncol*, 14, 391–436.
17. Gujarati, D. y Porter, D. (2010). *Econometría*. Santa Fe, México: McGraw-Hill.
18. Guzman, J. (2014). *Pronóstico de demanda y precios de energía en un mercado eléctrico*. (tesis de maestría). Universidad Michoacana de San Nicolás de Hidalgo, Michoacán, México.
19. Haykin, S. (1999). *Neural Network: A comprehensive foundation*. India: Pearson.
20. Hanke, J. y Wichern, D. (2010). *Pronósticos en los negocios*. México: Pearson.
21. Khare, M. y Nagendra, S. (2007). *Artificial neural networks in vehicular pollution modelling*. EE.UU: Springer.
22. Kozhan, R. (2010). *Financial Econometrics - with EViews*, Dinamarca: Ventus.

23. Mercado, D., Pedraza, L. y Martínez, E. (2015). Comparison of Neural Network applied to prediction of times Series, Prospect, 13(2), 88-95.
24. Ministerio de Salud (2013). Análisis de la situación del cáncer en el Perú. Recuperado de <http://www.dge.gob.pe>
25. Novalés, A. (2000). Econometría, Madrid, España: McGraw-Hill.
26. Organización mundial de la salud (2012). Globocan 2012: Cancer Incidence, Mortality and Prevalence Worldwide. EU.: Organización mundial de la salud. Recuperado de <http://globocan.iarc.fr>
27. Ortiz, C. (2012). Modelos Econométricos y de Redes Neuronales para predecir la Oferta Maderera en México: ARIMA vs NAR y ARMAX vs NARX (tesis de maestría). Universidad Nacional Autónoma de México, Coyoacán, México.
28. Prado, R y West, M. (2010). Time Series Modeling, Computation, and Inference. EE.UU: CHAPMAN & HALL/CRC.
29. Roth, B. (2010). Pautas en ONCOLOGIA Diagnóstico, tratamiento y seguimiento del cáncer. Argentina: Instituto “Angel H. Roffo”.
30. Serrano, A., Soria, E., y Martín, J. (2009-2010). Open courseware. Redes neuronales artificiales. Llevado a cabo en el tercer ciclo del programa de doctorado del departamento de ingeniería electrónica. Valencia, España.
31. Shumway, R y Stoffer, D. (2011). Time Series Analysis and Its Applications with R Examples, EE.UU: Springer.
32. Siegel, S., y Castellan, J. (1998). Estadística no paramétrica: aplicada a las ciencias de la conducta, Mexico: Trillas.
33. Siegel, R., Miller, K., y, Jemal, A. (2017). Cancer Statistics, 2017, CA: A Cancer Journal for Clinicians, 67, 7-30.

34. Villavicencio, J. (2007). Introducción a series de tiempo. Recuperado de [http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4\\_BxecUaZmg%3D](http://www.estadisticas.gobierno.pr/iepr/LinkClick.aspx?fileticket=4_BxecUaZmg%3D)
35. Uriel, E Perió, A. (2000). Introducción al análisis de series temporales. Madrid, España: Editorial AC.
36. Wei, W. (2006). Time series analysis: Univariate and multivariate methods. EE.UU: Pearson.

## APÉNDICE Y ANEXOS

### ANEXO 1

*Tabla 48.* Correlograma del número de pacientes con cáncer en el HNAAA. Chiclayo 2006 - 2017

LAG	AC	PAC	Q	Prob>Q
1	0.618	0.618	56.15	0.000
2	0.638	0.414	116.33	0.000
3	0.580	0.181	166.42	0.000
4	0.579	0.159	216.78	0.000
5	0.553	0.102	263.05	0.000
6	0.560	0.116	310.77	0.000
7	0.509	0.008	350.57	0.000
8	0.478	-0.030	385.87	0.000
9	0.450	-0.022	417.35	0.000
10	0.428	-0.014	446.13	0.000
11	0.427	0.034	474.99	0.000
12	0.391	-0.019	499.34	0.000
13	0.323	-0.118	516.06	0.000
14	0.306	-0.049	531.24	0.000
15	0.300	0.021	545.86	0.000
16	0.283	0.016	559.04	0.000
17	0.269	0.011	571.04	0.000
18	0.215	-0.067	578.75	0.000
19	0.189	-0.038	584.77	0.000
20	0.172	-0.003	589.81	0.000
21	0.175	0.035	595.06	0.000
22	0.128	-0.054	597.89	0.000
23	0.100	-0.067	599.63	0.000
24	0.071	-0.027	600.51	0.000
25	0.024	-0.068	600.61	0.000
26	0.020	-0.025	600.68	0.000
27	-0.032	-0.094	600.86	0.000
28	-0.054	-0.062	601.38	0.000
29	-0.117	-0.098	603.89	0.000
30	-0.121	-0.015	606.60	0.000
31	-0.179	-0.083	612.58	0.000
32	-0.163	-0.003	617.58	0.000
33	-0.153	0.100	621.98	0.000
34	-0.177	0.052	627.97	0.000
35	-0.205	0.001	636.05	0.000
36	-0.205	0.019	644.21	0.000
37	-0.213	0.022	653.13	0.000
38	-0.225	-0.010	663.16	0.000

LAG	AC	PAC	Q	Prob>Q
39	-0.240	-0.026	674.70	0.000
40	-0.265	-0.060	688.88	0.000
41	-0.223	0.070	698.98	0.000
42	-0.213	0.080	708.35	0.000
43	-0.245	-0.047	720.84	0.000
44	-0.213	0.021	730.42	0.000
45	-0.255	-0.067	744.26	0.000
46	-0.270	-0.052	759.85	0.000
47	-0.262	-0.017	774.71	0.000
48	-0.224	0.068	785.69	0.000
49	-0.244	-0.010	798.88	0.000
50	-0.239	-0.011	811.67	0.000
51	-0.223	0.065	822.96	0.000
52	-0.237	-0.007	835.81	0.000
53	-0.192	0.043	844.31	0.000
54	-0.173	0.047	851.27	0.000
55	-0.131	0.115	855.36	0.000
56	-0.173	-0.068	862.47	0.000
57	-0.165	-0.074	869.02	0.000
58	-0.103	0.062	871.63	0.000
59	-0.104	-0.026	874.28	0.000
60	-0.036	0.092	874.60	0.000
61	-0.069	-0.042	875.82	0.000
62	-0.043	-0.021	876.29	0.000
63	-0.027	0.021	876.48	0.000
64	0.003	0.047	876.48	0.000



## ANEXO 2



"Decenio de la Igualdad de oportunidades para mujeres y hombres"  
"Año del Diálogo y la Reconciliación Nacional"

### CENTRO DE INVESTIGACIÓN RED ASISTENCIAL DE LAMBAYEQUE

## CONSTANCIA N° 63

EL PRESIDENTE DEL CENTRO DE INVESTIGACIÓN DEL HOSPITAL  
NACIONAL ALMANZOR AGUINAGA ASENJO – RED ASISTENCIAL  
LAMBAYEQUE – ESSALUD, deja constancia

NIT: 1298 – 2018 – 3789

Que, la Gerencia de la Red Asistencial Lambayeque, autoriza la ejecución del Proyecto de Investigación que se detalla, el mismo que ha sido revisado y aprobado por el Comité de Investigación y el Comité Institucional de Ética en Investigación de la Red Asistencial Lambayeque "Juan Aita Valle" – EsSalud, de acuerdo con la Directiva N° 04-IETSI-ESSALUD. Directiva que establece los Lineamientos de Regulación y Fomento de la Investigación en EsSalud.

TITULO	"MODELO DE SERIE DE TIEMPO PARA PRONOSTICAR EL NUMERO DE PACIENTES CON CANCER, HOSPITAL NACIONAL ALMANZOR AGUINAGA ASENJO PERIODO 1997 - 2017"
AUTOR (ES)	• BACH. ESTAD. PERCY ANTONY QUISPE CHANTA
ASESOR (ES)	• Dr. ALFONSO TESEN ARROYO
UNIVERSIDAD	UNIVERSIDAD NACIONAL PEDRO RUIZ GALLO
ESCUELA	PROFESIONAL DE ESTADISTICA
FACULTAD	CIENCIAS FISICAS Y MATEMATICA
MOTIVO DE LA INVESTIGACION	Optar título de Licenciado en Estadística
AREA	SERVICIO DE ONCOLOGIA MEDICA y ARCHIVO CLINICO del HOSPITAL NACIONAL ALMANZOR AGUINAGA ASENJO de la RED ASISTENCIAL DE LAMBAYEQUE
HORARIO	A determinar por las Jefaturas antes indicadas
PERIODO	Del 28 de Junio al 28 de Setiembre 2018

Así mismo, el autor del mencionado Proyecto alcanzará al Área de Biblioteca – Red Asistencial Lambayeque, un ejemplar de la Tesis presentada.

Se expide la presente, para fines convenientes, ante las instancias correspondientes.

Chiclayo, 28 de Junio del 2018

CCR/cva

Dr. Carlos Isaias Carrón Rivera  
Presidente del Centro de Investigación  
Hospital Nacional Almanzor Aguinaga Asenjo



[centroinvest.lambayeque@essalud.gob.pe](mailto:centroinvest.lambayeque@essalud.gob.pe)

Plaza de la Seguridad Social S/N  
Chiclayo – Lambayeque – Perú  
Telf. (074) 237776 Anexo 2375  
(074) 221555

### ANEXO 3

#### Prueba de estacionariedad para la serie original

##### Hipótesis:

$H_0$ : La serie original no es estacionaria: Tiene raíz unitaria.

$H_1$ : La serie original es estacionaria: No tiene raíz unitaria.

##### Significancia:

$$\alpha = 0.05$$

##### Solución:

Dickey-Fuller	Lag order	p-value
-1.8425	5	0.642

##### Decisión:

Se acepta  $H_0$ .

##### Conclusión:

Dado que el p-valor = 0.642 > 0.05, se concluye que la serie original no es estacionaria.

## ANEXO 4

### Prueba de estacionariedad para la serie diferenciada

#### Hipótesis:

$H_0$ : La serie diferenciada en un periodo no es estacionaria: Tiene raíz unitaria.

$H_1$ : La serie diferenciada en un periodo es estacionaria: No tiene raíz unitaria.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Dickey-Fuller	Lag order	p-value
-6.782	5	0.01

#### Decisión:

Se rechaza  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.01 < 0.05, se concluye que la serie diferenciada es estacionaria.

## ANEXO 5

### Prueba de Ljung-Box de los residuos del modelo $ARIMA(1, 1, 0)$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $ARIMA(1,1,0)$  se distribuyen de forma independiente.

$H_1$ : Los residuos del modelo  $ARIMA(1,1,0)$  no se distribuyen de forma independiente.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Chi-Cuadrado	GL	P-Valor
2.828	1	0.093

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.093 > 0.05, se concluye que los residuos del modelo  $ARIMA(1,1,0)$  se distribuyen de forma independiente.

## ANEXO 6

### Prueba ANOVA de los residuos del modelo $ARIMA(1, 1, 0)$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $ARIMA(1,1,0)$  tienen promedio igual a cero.

$H_1$ : Los residuos del modelo  $ARIMA(1,1,0)$  tienen promedio diferente a cero.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	3	2.6	0.005	0.942
Residuals	142	68751	484.2		

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.942 > 0.05, se concluye que los residuos del modelo  $ARIMA(1,1,0)$  tienen promedio igual a cero.

## ANEXO 7

### Prueba Levene de los residuos del modelo $ARIMA(1, 1, 0)$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $ARIMA(1,1,0)$  tienen varianzas homogéneas.

$H_1$ : Los residuos del modelo  $ARIMA(1,1,0)$  no tienen varianzas homogéneas.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Df	Df	F value	Pr(>F)
11	132	1.664	0.089

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.089 > 0.05, se concluye que los residuos del modelo  $ARIMA(1,1,0)$  tienen varianzas homogéneas.

## ANEXO 8

### Prueba de Kolmogorov-Smirnov de los residuos del modelo $ARIMA(1, 1, 0)$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $ARIMA(1,1,0)$  siguen una distribución normal.

$H_1$ : Los residuos del modelo  $ARIMA(1,1,0)$  no siguen una distribución normal.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.035	0.995

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.995 > 0.05, se concluye que los residuos del modelo  $ARIMA(1,1,0)$  siguen una distribución normal.

## ANEXO 9

### Prueba de Ljung-Box de los residuos del modelo $ARIMA(2, 1, 0)$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $ARIMA(2,1,0)$  se distribuyen de forma independiente.

$H_1$ : Los residuos del modelo  $ARIMA(2,1,0)$  no se distribuyen de forma independiente.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Chi-Cuadrado	GL	P-Valor
0.492	1	0.483

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.483 > 0.05, se concluye que los residuos del modelo  $ARIMA(2,1,0)$  se distribuyen de forma independiente.



## ANEXO 10

### Prueba ANOVA de los residuos del modelo $ARIMA(2, 1, 0)$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $ARIMA(2,1,0)$  tienen promedio igual a cero.

$H_1$ : Los residuos del modelo  $ARIMA(2,1,0)$  tienen promedio diferente a cero.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	22	22.1	0.051	0.821
Residuals	142	61443	432.7		

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.821 > 0.05, se concluye que los residuos del modelo  $ARIMA(2,1,0)$  tienen promedio igual a cero.

## ANEXO 11

### Prueba Levene de los residuos del modelo $ARIMA(2, 1, 0)$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $ARIMA(2,1,0)$  tienen varianzas homogéneas.

$H_1$ : Los residuos del modelo  $ARIMA(2,1,0)$  no tienen varianzas homogéneas.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Df	F value	Pr(>F)
	11	132	1.713	0.077

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.077 > 0.05, se concluye que los residuos del modelo  $ARIMA(2,1,0)$  tienen varianzas homogéneas.

## ANEXO 12

### Prueba de Kolmogorov-Smirnov de los residuos del modelo $ARIMA(2, 1, 0)$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $ARIMA(2,1,0)$  siguen una distribución normal.

$H_1$ : Los residuos del modelo  $ARIMA(2,1,0)$  no siguen una distribución normal.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.046	0.915

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.915 > 0.05, se concluye que los residuos del modelo  $ARIMA(2,1,0)$  siguen una distribución normal.

## ANEXO 13

### Prueba de Ljung-Box de los residuos del modelo $ARIMA(3, 1, 0)$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $ARIMA(3,1,0)$  se distribuyen de forma independiente.

$H_1$ : Los residuos del modelo  $ARIMA(3,1,0)$  no se distribuyen de forma independiente.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Chi-Cuadrado	GL	P-Valor
0.145	1	0.704

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.704 > 0.05, se concluye que los residuos del modelo  $ARIMA(3,1,0)$  se distribuyen de forma independiente.

## ANEXO 14

### Prueba ANOVA de los residuos del modelo $ARIMA(3, 1, 0)$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $ARIMA(3,1,0)$  tienen promedio igual a cero.

$H_1$ : Los residuos del modelo  $ARIMA(3,1,0)$  tienen promedio diferente a cero.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	11	11.1	0.017	0.896
Residuals	142	90981	640.7		

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.896 > 0.05, se concluye que los residuos del modelo  $ARIMA(3,1,0)$  tienen promedio igual a cero.

## ANEXO 15

### Prueba Levene de los residuos del modelo $ARIMA(3, 1, 0)$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $ARIMA(3,1,0)$  tienen varianzas homogéneas.

$H_1$ : Los residuos del modelo  $ARIMA(3,1,0)$  no tienen varianzas homogéneas.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Df	F value	Pr(>F)
	11	132	1.669	0.087

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.087 > 0.05, se concluye que los residuos del modelo  $ARIMA(3,1,0)$  tienen varianzas homogéneas.

## ANEXO 16

### Prueba de Kolmogorov-Smirnov de los residuos del modelo $ARIMA(3, 1, 0)$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $ARIMA(3,1,0)$  siguen una distribución normal.

$H_1$ : Los residuos del modelo  $ARIMA(3,1,0)$  no siguen una distribución normal.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.041	0.966

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.966 > 0.05, se concluye que los residuos del modelo  $ARIMA(3,1,0)$  siguen una distribución normal.

## ANEXO 17

### Prueba de Ljung-Box de los residuos del modelo $SARIMA(1, 1, 0)(1, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  se distribuyen de forma independiente.

$H_1$ : Los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  no se distribuyen de forma independiente.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Chi-Cuadrado	GL	P-Valor
2.310	1	0.129

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.129 > 0.05, se concluye que los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  se distribuyen de forma independiente.



## ANEXO 18

### Prueba ANOVA de los residuos del modelo $SARIMA(1, 1, 0)(1, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(1, 1, 0)(1, 1, 0)_{12}$  tienen promedio igual a cero.

$H_1$ : Los residuos del modelo  $SARIMA(1, 1, 0)(1, 1, 0)_{12}$  tienen promedio diferente a cero.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	11	11.1	0.017	0.896
Residuals	142	90981	640.7		

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.896 > 0.05, se concluye que los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  tienen promedio igual a cero.

## ANEXO 19

### Prueba Levene de los residuos del modelo $SARIMA(1, 1, 0)(1, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  tienen varianzas homogéneas.

$H_1$ : Los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  no tienen varianzas homogéneas.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Df	Df	F value	Pr(>F)
11	132	2.977	0.001

#### Decisión:

Se rechaza  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.001 < 0.05, se concluye que los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  no tienen varianzas homogéneas.

## ANEXO 20

### Prueba de Kolmogorov-Smirnov de los residuos del modelo $SARIMA(1, 1, 0)(1, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  siguen una distribución normal.

$H_1$ : Los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  no siguen una distribución normal.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.074	0.406

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.406 > 0.05, se concluye que los residuos del modelo  $SARIMA(1,1,0)(1,1,0)_{12}$  siguen una distribución normal.

## ANEXO 21

### Prueba de Ljung-Box de los residuos del modelo $SARIMA(1, 1, 0)(2, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  se distribuyen de forma independiente.

$H_1$ : Los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  no se distribuyen de forma independiente.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Chi-Cuadrado	GL	P-Valor
2.372	1	0.124

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.124 > 0.05, se concluye que los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  se distribuyen de forma independiente.

## ANEXO 22

### Prueba ANOVA de los residuos del modelo $SARIMA(1, 1, 0)(2, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(1, 1, 0)(2, 1, 0)_{12}$  tienen promedio igual a cero.

$H_1$ : Los residuos del modelo  $SARIMA(1, 1, 0)(2, 1, 0)_{12}$  tienen promedio diferente a cero.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	12	12.2	0.021	0.885
Residuals	142	82545	581.3		

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.885 > 0.05, se concluye que los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  tienen promedio igual a cero.

## ANEXO 23

### Prueba Levene de los residuos del modelo $SARIMA(1, 1, 0)(2, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  tienen varianzas homogéneas.

$H_1$ : Los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  no tienen varianzas homogéneas.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Df	F value	Pr(>F)
	11	132	2.593	0.005

#### Decisión:

Se rechaza  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.005 < 0.05, se concluye que los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  no tienen varianzas homogéneas.

## ANEXO 24

### Prueba de Kolmogorov-Smirnov de los residuos del modelo $SARIMA(1, 1, 0)(2, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  siguen una distribución normal.

$H_1$ : Los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  no siguen una distribución normal.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.080	0.321

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.321 > 0.05, se concluye que los residuos del modelo  $SARIMA(1,1,0)(2,1,0)_{12}$  siguen una distribución normal.

## ANEXO 25

### Prueba de Ljung-Box de los residuos del modelo $SARIMA(2, 1, 0)(1, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  se distribuyen de forma independiente.

$H_1$ : Los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  no se distribuyen de forma independiente.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Chi-Cuadrado	GL	P-Valor
0.488	1	0.485

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.485 > 0.05, se concluye que los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  se distribuyen de forma independiente.



## ANEXO 26

### Prueba ANOVA de los residuos del modelo $SARIMA(2, 1, 0)(1, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(2, 1, 0)(1, 1, 0)_{12}$  tienen promedio igual a cero.

$H_1$ : Los residuos del modelo  $SARIMA(2, 1, 0)(1, 1, 0)_{12}$  tienen promedio diferente a cero.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	0	0.1	0.000	0.989
Residuals	142	86003	605.7		

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.989 > 0.05, se concluye que los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  tienen promedio igual a cero.

## ANEXO 27

### Prueba Levene de los residuos del modelo $SARIMA(2, 1, 0)(1, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  tienen varianzas homogéneas.

$H_1$ : Los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  no tienen varianzas homogéneas.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Df	F value	Pr(>F)
	11	132	0.799	0.641

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.641 > 0.05, se concluye que los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  tienen varianzas homogéneas.

## ANEXO 28

### Prueba de Kolmogorov-Smirnov de los residuos del modelo $SARIMA(2, 1, 0)(1, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  siguen una distribución normal.

$H_1$ : Los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  no siguen una distribución normal.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.080	0.321

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.321 > 0.05, se concluye que los residuos del modelo  $SARIMA(2,1,0)(1,1,0)_{12}$  siguen una distribución normal.

## ANEXO 29

### Prueba de Ljung-Box de los residuos del modelo $SARIMA(2, 1, 0)(2, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  se distribuyen de forma independiente.

$H_1$ : Los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  no se distribuyen de forma independiente.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Chi-Cuadrado	GL	P-Valor
0.531	1	0.466

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.466 > 0.05, se concluye que los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  se distribuyen de forma independiente.

### ANEXO 30

#### Prueba ANOVA de los residuos del modelo $SARIMA(2, 1, 0)(2, 1, 0)_{12}$

##### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  tienen promedio igual a cero.

$H_1$ : Los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  tienen promedio diferente a cero.

##### Significancia:

$$\alpha = 0.05$$

##### Solución:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	20	20.3	0.037	0.848
Residuals	142	77899	548.6		

##### Decisión:

Se acepta  $H_0$ .

##### Conclusión:

Dado que el p-valor = 0.848 > 0.05, se concluye que los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  tienen promedio igual a cero.

## ANEXO 31

### Prueba Levene de los residuos del modelo $SARIMA(2, 1, 0)(2, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  tienen varianzas homogéneas.

$H_1$ : Los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  no tienen varianzas homogéneas.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Df	F value	Pr(>F)
	11	132	2.419	0.009

#### Decisión:

Se rechaza  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.009 < 0.05, se concluye que los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  no tienen varianzas homogéneas.

## ANEXO 32

### Prueba de Kolmogorov-Smirnov de los residuos del modelo $SARIMA(2, 1, 0)(2, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  siguen una distribución normal.

$H_1$ : Los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  no siguen una distribución normal.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.077	0.361

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.361 > 0.05, se concluye que los residuos del modelo  $SARIMA(2,1,0)(2,1,0)_{12}$  siguen una distribución normal.

### ANEXO 33

#### Prueba de Ljung-Box de los residuos del modelo $SARIMA(3, 1, 0)(1, 1, 0)_{12}$

##### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  se distribuyen de forma independiente.

$H_1$ : Los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  no se distribuyen de forma independiente.

##### Significancia:

$$\alpha = 0.05$$

##### Solución:

Chi-Cuadrado	GL	P-Valor
0.234	1	0.629

##### Decisión:

Se acepta  $H_0$ .

##### Conclusión:

Dado que el p-valor = 0.629 > 0.05, se concluye que los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  se distribuyen de forma independiente.



## ANEXO 34

### Prueba ANOVA de los residuos del modelo $SARIMA(3, 1, 0)(1, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  tienen promedio igual a cero.

$H_1$ : Los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  tienen promedio diferente a cero.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	34	34.1	0.060	0.807
Residuals	142	81172	571.6		

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.807 > 0.05, se concluye que los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  tienen promedio igual a cero.

## ANEXO 35

### Prueba Levene de los residuos del modelo $SARIMA(3, 1, 0)(1, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  tienen varianzas homogéneas.

$H_1$ : Los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  no tienen varianzas homogéneas.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Df	F value	Pr(>F)
	11	132	2.518	0.007

#### Decisión:

Se rechaza  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.007 < 0.05, se concluye que los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  no tienen varianzas homogéneas.

## ANEXO 36

### Prueba de Kolmogorov-Smirnov de los residuos del modelo $SARIMA(3, 1, 0)(1, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  siguen una distribución normal.

$H_1$ : Los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  no siguen una distribución normal.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.065	0.575

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.575 > 0.05, se concluye que los residuos del modelo  $SARIMA(3,1,0)(1,1,0)_{12}$  siguen una distribución normal.

## ANEXO 37

### Prueba de Ljung-Box de los residuos del modelo $SARIMA(3, 1, 0)(2, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  se distribuyen de forma independiente.

$H_1$ : Los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  no se distribuyen de forma independiente.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Chi-Cuadrado	GL	P-Valor
0.205	1	0.651

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.651 > 0.05, se concluye que los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  se distribuyen de forma independiente.

## ANEXO 38

### Prueba ANOVA de los residuos del modelo $SARIMA(3, 1, 0)(2, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  tienen promedio igual a cero.

$H_1$ : Los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  tienen promedio diferente a cero.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
residuales\$año	1	41	40.8	0.079	0.779
Residuals	142	73070	514.6		

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.779 > 0.05, se concluye que los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  tienen promedio igual a cero.

### ANEXO 39

#### Prueba Levene de los residuos del modelo $SARIMA(3, 1, 0)(2, 1, 0)_{12}$

##### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  tienen varianzas homogéneas.

$H_1$ : Los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  no tienen varianzas homogéneas.

##### Significancia:

$$\alpha = 0.05$$

##### Solución:

Df	Df	F value	Pr(>F)
11	132	2.186	0.019

##### Decisión:

Se rechaza  $H_0$ .

##### Conclusión:

Dado que el p-valor = 0.019 < 0.05, se concluye que los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  no tienen varianzas homogéneas.

## ANEXO 40

### Prueba de Kolmogorov-Smirnov de los residuos del modelo $SARIMA(3, 1, 0)(2, 1, 0)_{12}$

#### Hipótesis:

$H_0$ : Los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  siguen una distribución normal.

$H_1$ : Los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  no siguen una distribución normal.

#### Significancia:

$$\alpha = 0.05$$

#### Solución:

Prueba de normalidad	Kolmogorov-Smirnov	
	Estadístico	Sig
Residuos	0.066	0.554

#### Decisión:

Se acepta  $H_0$ .

#### Conclusión:

Dado que el p-valor = 0.554 > 0.05, se concluye que los residuos del modelo  $SARIMA(3,1,0)(2,1,0)_{12}$  siguen una distribución normal.