

**UNIVERSIDAD DE EL SALVADOR
FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE
DEPARTAMENTO DE MATEMÁTICA**



TRABAJO DE INVESTIGACION TITULADO:

Análisis de Series Temporales utilizando Modelos Arima y su Aplicación en la Predicción de Indicadores del Ministerio de Salud en el Departamento de Santa Ana

DOCENTE DIRECTOR:

LICDO. LUIS ERNESTO AQUINO SALAZAR

TRABAJO DE INVESTIGACIÓN PRESENTADO POR:

JAIME ISAAC PEÑA MEJÍA

PARA OPTAR AL GRADO DE:

LICENCIADO EN ESTADÍSTICA

DICIEMBRE 2014

SANTA ANA, EL SALVADOR, CENTRO AMÉRICA

UNIVERSIDAD DE EL SALVADOR

RECTOR

ING. MARIO ROBERTO NIETO LOVO

VICE-RECTOR ACADÉMICO

MSD. ANA MARÍA GLOWER DE ALVARADO

SECRETARIO GENERAL

DRA. ANA LETICIA ZA VALETA DE AMAYA

FISCAL GENERAL

LICDO. FRANCISCO CRUZ LETONA

FACULTAD MULTIDISCIPLINARIA DE OCCIDENTE

DECANO

LICDO. RAÚL ERNESTO AZCÚNAGA LÓPEZ

VICE-DECANO

ING. WILLIAM VIRGILIO ZAMORA GIRÓN

SECRETARIO DE LA FACULTAD

LICDO. VICTOR HUGO MERINO QUEZADA

JEFE DEL DEPARTAMENTO DE MATEMÁTICA

LICDA. SARA VILMA RODRÍGUEZ DE CHICAS

AGRADECIMIENTOS

A DIOS: Por estar siempre conmigo e iluminarme en el camino de la vida.

A MIS PADRES: Por su apoyo incondicional y por hacer oraciones que me ayudan e iluminan en el largo viaje del conocimiento y de la vida.

MI ASESOR: Por orientarme con sus conocimientos y consejos en la tesis.

AGRADECIMIENTOS ESPECIALES

Como investigador presento agradecimientos especiales, a las autoridades de SIBASI Santa Ana y a todo el personal que ahí labora, por haber proporcionado siempre una atenta y efectiva colaboración durante toda la investigación desarrollada en la institución, pues se nos facilitó las bases de datos necesarias para culminar con éxito esta tesis; además por demostrar en todo momento un espíritu de servicio y calidez humana.

Índice general

Índice general	II
Índice de figuras	VII
Índice de Tablas	X
Introducción	1
1 El Problema	3
1.1 Planteamiento del problema	3
1.2 Justificación	5
1.3 Objetivos	6
1.3.1 Objetivo General	6
1.3.2 Objetivos Específicos	6
2 Sobre SIBASI Santa Ana	7
2.1 Historia de SIBASI	7
2.1.1 Ley de creación del Sistema Nacional de Salud	7
2.2 Obtención de los datos	16

3 Modelos ARIMA	20
3.1 Conceptos Básicos	20
3.1.1 Estacionariedad	21
3.1.2 Tendencia	21
3.1.3 Estacionalidad	21
3.1.4 Función de Autocorrelación(FAS ó FAC)	22
3.1.5 Función de Autocorrelación Parcial(PACF ó FAP)	22
3.2 Principales modelos ARIMA de procesos de series temporales .	27
3.2.1 Introducción	27
3.2.2 Metodología Box-Jenkins	28
3.2.3 Modelos Autorregresivos(AR)	29
3.2.4 Modelos de Medias Móviles(MA)	30
3.2.5 Modelos de promedio móvil autorregresivos	31
3.3 Modelos Autoregresivo Integrados de Medias Móviles ARMA(p,d,q)	32
3.4 Proceso Estacional Autoregresivo Integrado y de Media Móvil	33
3.5 Criterios de Información	34
3.6 Estrategia de implementación de la construcción del modelo .	35
3.6.1 Paso 1: Identificación del modelo	36
3.6.2 Paso 2: Estimación del modelo	38

3.6.3	Paso 3: Verificación del modelo(Diagnosis)	39
3.6.4	Paso 4: Elaboración de pronósticos	41
3.7	Metodología	42
3.7.1	Tipo de estudio	42
3.7.2	Población	42
3.7.3	Muestra	42
3.7.4	Instrumentos	42
3.7.5	Procedimiento	42
4	Ajuste del Modelo	44
4.1	Identificación del modelo de la serie IRAS	44
4.1.1	Estacionariedad en media	44
4.1.2	Estacionariedad en varianza de IRAS	46
4.1.3	Estacionalidad IRAS	47
4.2	Estimación para el modelo IRAS	50
4.3	Diagnosis o Chequeo de IRAS	50
4.3.1	Incorrelación de residuos de IRAS	51
4.3.2	Normalidad de los residuos para IRAS	52
4.3.3	Predicciones IRAS	53
4.4	Identificación del modelo para Diarrea	56

4.4.1	Estacionariedad en media de Diarrea	56
4.4.2	Estacionariedad en varianza de Diarrea	57
4.4.3	Estacionalidad de la serie Diarrea	59
4.5	Estimación de la serie Diarrea.	62
4.6	Diagnosís o Chequeo de Diarrea	64
4.6.1	Incorrelación de residuos de Diarrea	64
4.6.2	Normalidad de los residuos para Diarrea	65
4.6.3	Predicciones de la serie Diarrea	67
4.7	Identificación del modelo para Dengue	70
4.7.1	Estacionariedad en media para Dengue	70
4.7.2	Estacionariedad en varianza de Dengue	71
4.7.3	Estacionalidad de la serie Dengue	73
4.8	Estimación Dengue	76
4.9	Diagnosís o Chequeo de Dengue	77
4.9.1	Incorrelación de los residuos de Dengue	77
4.9.2	Normalidad de los residuos de Dengue	78
4.10	Análisis del segundo modelo de la serie Dengue	82
	Conclusiones	85

Recomendaciones	86
Bibliografía	87
ANEXOS	89

Índice de figuras

2.1	Gráfico de la serie de tiempo IRAS	17
2.2	Gráfico de la serie de tiempo Dengue	18
2.3	Gráfico de la serie Diarrea	19
4.1	Gráfico de la serie de tiempo IRAS desde Enero 2007 hasta Diciembre 2013	44
4.2	Correlograma de la serie IRAS sin transformaciones	45
4.3	Gráfico de rango-media para IRAS sin transformación	46
4.4	Gráfico de IRAS con logaritmos	47
4.5	Correlograma de la serie IRAS con transformación logarítmica	48
4.6	Correlograma de la serie IRAS con logaritmos y una diferencia estacional	49
4.7	Correlograma de residuos	51
4.8	Gráfico QQ de residuos de la serie IRAS.	53
4.9	Predicciones IRAS.	54
4.10	Gráfico de la serie Diarrea sin transformaciones desde Enero 2007 hasta Diciembre 2013.	56

4.11	Correlograma de la serie Diarrea sin transformaciones	57
4.12	Gráfico de rango-media de la serie Diarrea	58
4.13	Gráfico de la serie Diarrea con transformación logarítmica . . .	59
4.14	Correlograma de la serie Diarrea con logaritmos.	59
4.15	Gráfico de la serie Diarrea con logaritmos y una diferencia estacional.	60
4.16	correlograma de la serie Diarrea con logaritmos y una diferen- cia estacional	61
4.17	Correlograma de residuos de la serie Diarrea.	64
4.18	Gráfico QQ de residuos de la serie Diarrea.	66
4.19	Gráfico de la serie de tiempo Diarrea y sus Predicciones	67
4.20	Gráfico de la serie de tiempo Dengue sin transformaciones des- de Enero 2007 a Diciembre 2013.	70
4.21	Correlograma de la serie Dengue con transformaciones.	71
4.22	Gráfico de la serie Dengue con logaritmos , Enero 2007-Diciembre 2013.	72
4.23	Gráfico de rango-media de la serie Dengue	73
4.24	Correlograma de la serie Dengue con logaritmos	74
4.25	Gráfico de la serie Dengue con logaritmos y una diferencia estacional.	75

4.26 Correlograma de la serie Dengue con logaritmos y una diferencia estacional.	75
4.27 Correlograma de los residuos de la serie Dengue	77
4.28 Gráfico QQ de los residuos de la serie Dengue	79
4.29 Gráfico de la serie y pronósticos de serie Dengue	80
4.30 Correlograma de residuos para el segundo modelo de Dengue .	83

Índice de Tablas

4.1	Datos y Predicciones de casos de IRAS para los años 2013-2014	54
4.3	Datos y Predicciones de casos de Diarrea para los años 2013-2014	67
4.4	Datos y pronósticos de la serie Dengue para los años 2013-2014	80
4.5	Anexo 1: Datos de las series temporales IRAS, Diarrea y Dengue.	90
4.6	ANEXO 2: Datos y predicciones de casos de IRAS mensual	92
4.11	ANEXO 3: Datos y Predicciones de casos de <i>Diarrea</i> .	97
4.16	Anexo 4: Datos y Pronósticos para la serie Dengue. . .	102

Introducción

La importancia de saber el comportamiento de determinada enfermedad en un futuro permite tomar previsiones para evitar tomar riesgos innecesarios o tener la oportunidad de obtener mayores beneficios que la mayoría de las personas no puede aprovechar. Las técnicas de pronóstico de series de tiempo por los métodos estadísticos tradicionales permiten llevar a cabo esta labor como, por ejemplo, por medio de la metodología Box Jenkins, la cual permite obtener buenas aproximaciones en el caso de que dicho método sea bien aplicado, ya que ésta metodología tiene una parte de arte en el sentido de que el proceso es iterativo hasta alcanzar el mejor modelo de acuerdo a la experiencia del investigador. Los modelos de series de tiempo han sido tradicionalmente aplicados a *Ciencias de la Salud*, ya que estos han probado en diversos estudios ser efectivos en la predicción. Por lo que, el modelo ARIMA(modelos Autorregresivos Integrados de Medias Móviles) es ampliamente reconocido en estudios de salud, ya que han permitido modelar con algunas técnicas adicionales(modelos heteroscedásticos) la complejidad de las series . El modelo antes mencionado pertenece a la clasificación de modelos paramétricos, ya que parten de una función de distribución, y reducen el problema a estimar los parámetros que mejor se ajustan a las observaciones de la muestra¹. Dichos modelos resultan ser muy potentes cuando el proceso generador de datos sigue una distribución propuesta, en la teoría los resultados más conocidos se han obtenido a través de la distribución normal o

¹Bonilla,2003

gaussiana². Surge así la necesidad de contar con una herramienta estadística para abordar el modelado de Series de tiempo para pronosticar números de casos de Infecciones Respiratorias Agudas Superiores (IRAS), Dengue y Diarrea para el Departamento de Santa Ana. En este sentido; este trabajo pretende realizar un estudio comprensivo del análisis de Series Temporales con observaciones que representan número de casos de enfermedades. Los pronósticos se realizan utilizando la metodología ARIMA porque se ha demostrado que sus estimaciones son bastantes buenas en diversos estudios. De esta forma, presentamos las ideas esenciales que serán desarrolladas en este trabajo y que nos ayudarán no sólo a explorar un área más de las estadísticas contemporáneas, sino, que permitirán aclarar las ideas que no han podido explicarse por falta de teoría y herramientas que las refuercen.

²Venegas,2008

Capítulo 1

El Problema

1.1. Planteamiento del problema

En el pasado los estadísticos, han tratado con modelos ARIMA regulares y estacionarios o no estacionarios, para pronosticar en áreas como Economía, Ecología, Biología y Ciencias de la salud. En el área de salud en nuestro país casi no hay estudios que impliquen el uso de los modelos Autorregresivos de Medias Móviles, las técnicas clásicas de Modelización y Series Temporales pueden ser aplicadas debido a que se vuelve necesario tener una idea sobre el número de casos de algunas enfermedades en el Departamento de Santa Ana para concientizar a la población de tomar las medidas necesarias para mantener la salud pública. Por lo anterior se busca identificar un modelo ARIMA que permita predecir el número de casos de IRAS, Dengue y Diarrea para el año 2014 y así poder tener una muy buena idea de lo que ocurrirá en el futuro con las enfermedades antes mencionadas y poder tomar las decisiones más adecuadas para poder evitar que se den más casos de dichas enfermedades. Por lo tanto la idea central de este trabajo se focaliza en facilitar los fundamentos que permitan modelar y predecir una Serie Temporal con observaciones que representan número de enfermedades, sin dejar pasar la teoría de los modelos ARIMA que se estudia en el curso de series temporales. Durante el desarrollo de este trabajo, se dará respuesta a las siguientes interrogantes, que se originan a partir de la metodología Box-Jenkins.

¿Que es un modelo ARIMA ?

¿Por qué son importantes los modelos ARIMA?

¿Cuál es el número de casos de Dengue, IRAS y de Diarrea en cada mes del año 2014?

¿Cuáles son las herramientas estadísticas requeridas para poder realizar una aplicación de un Análisis de Series Temporales utilizando modelos ARIMA?

1.2. Justificación

Hacer un estudio sobre **Modelos Autorregresivos Integrados de Medias Móviles** para predecir el número de casos de Dengue, Infecciones Respiratorias Agudas y Diarrea en el Departamento de Santa Ana tiene sobradas justificaciones. En primer lugar es importante hacer aportes concretos que vayan en beneficio de la sociedad, sobre todo si es en áreas que inciden directamente en la población como el área de salud pública. Predecir el número de casos que se tendrán de las enfermedades antes mencionadas para el año 2014 juega un papel fundamental para el Ministerio de Salud y el desarrollo del país, porque les permitirá desarrollar estrategias que se fundamenten en reducir el número de enfermedades que tanto afectan a la población del Departamento de Santa Ana, la predicción del número de enfermedades utilizando modelos estadísticos como los Modelos Autorregresivos Integrados de Medias Móviles ha sido tratada de manera superficial o casi nula en El Salvador . Como sabemos la investigación es uno de los fines de la Universidad de El Salvador y es importante hacer encajar los estudios teóricos en la práctica y así darle un mayor sentido a los estudios que tienen relación con la vida práctica. En particular se considera que la teoría relativa a los modelos Autorregresivos Integrados de Medias Móviles en series temporales tiene importantes aplicaciones en la realidad, en particular en el área de salud, y esto es lo que se pretende mostrar con el desarrollo del presente trabajo. Bajo estas premisas y tomando en cuenta la importancia de tener un modelo de series temporales que se ajuste al comportamiento del número de casos de dengue, Infecciones Respiratorias Agudas y Diarrea en el Departamento de Santa Ana, además dado que no se ha hecho un estudio de tal envergadura en Santa Ana, queda justificado el desarrollo de la investigación.

1.3. Objetivos

1.3.1. Objetivo General

Proponer un modelo Autorregresivo Integrado de Medias Móviles que se ajuste a las series temporales Diarrea, Dengue e IRAS (Infecciones Respiratorias Agudas Superiores) para realizar pronósticos y así determinar el número de casos estimados para el Departamento de Santa Ana en el año 2014.

1.3.2. Objetivos Específicos

1. Identificar un modelo Autorregresivo Integrado de Medias Móviles aplicando la metodología descrita por Box-Jenkins a las series temporales Dengue, Diarrea e IRAS(Infecciones Respiratorias Agudas)
2. Pronosticar el número de casos de Diarrea, Dengue e IRAS aplicando la metodología descrita por Box - Jenkins

Capítulo 2

Sobre SIBASI Santa Ana

2.1. Historia de SIBASI

La Institución nace el 23 de Julio de 1900, con el nombre de Consejo Superior de Salubridad, dependencia del Ministerio de Gobernación. El Consejo determinó que entre las actividades principales a realizar fueran: estadísticas médicas, saneamiento de zonas urbanas, inspecciones de víveres, higiene de rastros y mercados, construcción de cloacas y sistemas de aguas servidas, obligatoriedad de instalar letrinas, lucha contra los mosquitos, visitas a establos, fábricas y beneficios de lavar café. El primer Código de Sanidad entra en vigencia el 24 de julio del mismo año. El 15 de agosto se creó la Dirección General de Vacunación la cual en 1907 se llamó Instituto de Vacunación Contra la Viruela³.

2.1.1. Ley de creación del Sistema Nacional de Salud

Decreto No. 442

Art. 1. Créase el Sistema Nacional de Salud de El Salvador, en adelante SNS o el Sistema, que estará constituido por un conjunto de instituciones que forman parte de la administración pública interrelacionadas e integradas en su funcionamiento, de manera armónica y sistematizada y cuya finalidad es elaborar y ejecutar políticas públicas que garanticen el derecho a la salud de la población.

³Diario Oficial del MINISTERIO de Salud, 2002

Decreto No. 77 Considerando:

- Que por Decreto Legislativo No. 775, de fecha 10 de agosto de 2005, publicado en el Diario Oficial No. 161, Tomo No. 368 del 1 de septiembre de ese mismo año, se emitió la Ley del Sistema Básico de Salud Integral;

En el Art.19 de la *Ley del SIBASI* se establece que, los Sistemas Básicos de Salud Integral dependerán jerárquicamente de la Dirección Regional correspondiente . Los Sistemas Básicos de Salud Integral se conformarán con base a la división política y administrativa del país, distribución poblacional por departamento y las necesidades sanitarias que la población demanda; los SIBASI están distribuidos uno por Departamento y tres para San Salvador. Las funciones de SIBASI las especifica el Art.20 de la Ley del SIBASI algunas son:

1. Elaborar, ejecutar y controlar el plan operativo del SIBASI con base a su diagnóstico de salud y recursos disponibles;
2. Verificar que la atención integral a las personas y al medio ambiente se proporcione con base a la normativa establecida;
3. Promover la participación de la población y los actores sociales para la conservación de la salud; la prevención del daño y sus complicaciones;
4. Apoyar a los establecimientos de la red, mediante el intercambio de recursos, el desarrollo de actividades de capacitación, supervisión y asistencia técnica;
5. Conducir y controlar los procesos de la provisión de los servicios de salud, así como el manejo

6. Desarrollar la vigilancia epidemiológica con la participación de los diferentes actores sociales y establecimientos de sector del área de su responsabilidad;

En el numeral 6 se encuentra la función de "Desarrollar la Vigilancia epidemiológica", que en parte se desarrolla bajo las directrices establecidas por el Art.133 del código de salud que establece las enfermedades que están bajo vigilancia. La vigilancia epidemiológica es un conjunto de actividades y procedimientos sobre enfermedades, muertes y síndromes sujetos a vigilancia y a la notificación obligatoria, que generan información sobre el comportamiento y la tendencia de los mismos, para la implementación de intervenciones en forma oportuna, a fin de lograr el control inmediato de dichos eventos⁴.

La vigilancia epidemiológica es, sin lugar a dudas, una de las bases fundamentales para cumplir, con eficacia, eficiencia y equidad, las demás tareas y actividades que realiza el equipo de salud en todos los niveles, especialmente en el nivel local.

La vigilancia epidemiológica no es un proceso adicional sino inherente a todas las acciones que buscan la prevención y el control de los problemas de salud, por lo tanto, todos(as) los(as) funcionario(as) de salud, en su quehacer diario, son partícipes de este proceso, ya sea generando información, analizándola, tomando decisiones o ejecutando las decisiones mediante la realización de las intervenciones de manera coordinada con el equipo de trabajo institucional, interinstitucional y con la comunidad.

Art. 132. Las enfermedades objeto del Reglamento Sanitario Internacional son: viruela, fiebre amarilla selvática y urbana, peste y cólera. La declaración de estas enfermedades es obligatoria en el término de veinticuatro horas

⁴Código de Salud, El Salvador

siguientes a su diagnóstico, sea este cierto o probable. Esta información deberá comunicarse al Ministerio o su dependencia más cercana.

En el código de salud de El Salvador, en su sección 23 *Enfermedades objeto de vigilancia*, en el Art.133 determina las enfermedades que estarán bajo vigilancia, en dicho artículo se hace referencia al Art.132 en donde se presenta de forma explícita el listado de enfermedades haciendo la aclaratoria que se pueden agregar o suprimir enfermedades según el ministerio estime conveniente. Las enfermedades que están bajo vigilancia epidemiológica se han clasificado en los siguientes grupos:

Grupo 1: Enfermedades Inmuno prevenibles

- Difteria
- Meningitis tuberculosa
- Parálisis flácida aguda
- Sarampión
- Tétanos neonatal
- Tétanos.
- Tosferina
- Hepatitis aguda tipo A
- Hepatitis aguda tipo B
- Parotiditis infecciosa
- Rubéola
- Rubéola congénita

- Varicela

Grupo 2: Enfermedades intestinales

- Cólera
- Fiebre tifoidea
- Diarrea, enteritis y gastroenteritis
- Amibiasis
- Giardiasis
- Helmintiasis

Grupo 3: Infecciones con modo de transmisión preferentemente sexual

- SIDA
- Infección por VIH
- Sífilis congénita
- Sífilis adquirida y no especificada
- Infección gonocócica del tracto genitourinario inferior.
- Herpes genital
- Condiloma acuminado
- Linfogranuloma venéreo

- Chancro blando
- Tricomoniasis urogenital
- Candidiasis de vulva y de vagina

Grupo 4: Enfermedades meníngeas

- Meningitis por Haemófilos
- Meningitis Meningocócica

Grupo 5: Otras de interés epidemiológico

- Tuberculosis pulmonar
- Influenza
- Neumonías
- Síndrome respiratorio agudo severo (SARS)
- Infección aguda de vías respiratorias superiores (IRAS)
- conjuntivitis hemorrágica
- Conjuntivitis bacteriana aguda
- Lepra
- Enfermedad vírica de Marburg y Ebola

Grupo 6: Enfermedades vectorizadas por artrópodos

- Dengue hemorrágico
- Dengue clásico
- Paludismo
- Fiebre amarilla
- Leishmaniasis
- Enfermedad de Chagas aguda
- Fiebre equina venezolana
- Encefalitis viral no especificada
- Tifus epidémico

Grupo 7:

- Brucelosis
- Carbunco (Ántrax)
- Leptospirosis
- Rabia humana
- Hanta virus

Grupo 8: Enfermedades crónicas degenerativas

- Cáncer Cervico-uterino
- Hipertensión arterial

- Insuficiencia renal crónica

Grupo 9: Enfermedades metabólicas y de la nutrición

- Diabetes mellitus
- Bocio endémico
- Desnutrición proteicocalórica leve
- Desnutrición proteicocalórica moderada
- Desnutrición proteicocalórica severa

Grupo 10: Intoxicaciones alimentarias

- Intoxicación alimentaria aguda

Grupo 11: Agresiones por animales

- Mordedura por animales transmisores de rabia
- Picadura por abeja africanizada
- Mordedura por serpiente venenosa

Grupo 12: Trastornos mentales y del comportamiento

- Intento de suicidio (conducta suicida)
- Drogodependencia
- Depresión

- Ansiedad

- Alcoholismo

Como puede observarse es extenso el listado de enfermedades que están bajo vigilancia epidemiológica, pero el presente estudio se hace únicamente para el número de casos registrados en el departamento de Santa Ana, correspondientes a las enfermedades Diarreicas, Dengue e IRAS⁵. Por lo que quedan muchas enfermedades no abordadas en este trabajo, para que en un futuro pueda realizarse un estudio similar con ellas.

Dichas enfermedades se seleccionaron debido a que actualmente la Diarrea, Dengue e IRAS son las enfermedades que presentan mayor número de casos. Esta selección se determinó en una entrevista con el Director de **SIBASI**, quien expresó que estas enfermedades son las más comunes en el Departamento de Santa Ana y están bajo vigilancia debido a que en años anteriores se han dado muchas defunciones de habitantes de Santa Ana, lo cual es una situación preocupante para el Ministerio de Salud. El Director de **SIBASI** expresa que no se descuide el implemento de las medidas de higiene para evitar un aumento de los casos de Infecciones Respiratorias Agudas Superiores (IRAS) y de Dengue, sobre todo porque los niños son los más afectados. El lavado de manos, cubrirse bien al toser y estornudar son parte de las medidas. Gracias a la coordinación y cooperación del departamento de epidemiología, estadística e informática, de SIBASI⁶ Santa Ana, se tuvo acceso a la base de datos de las enfermedades a las que se enfocará el estudio, donde las bases de datos de cada una de las enfermedades representa una serie de tiempo, y éstas observaciones han sido recolectadas mensualmente⁷, y el periodo

⁵Infecciones Respiratorias Agudas Superiores

⁶Sistemas Básicos de Salud Integral

⁷Se conoce como "Semana epidemiológica"

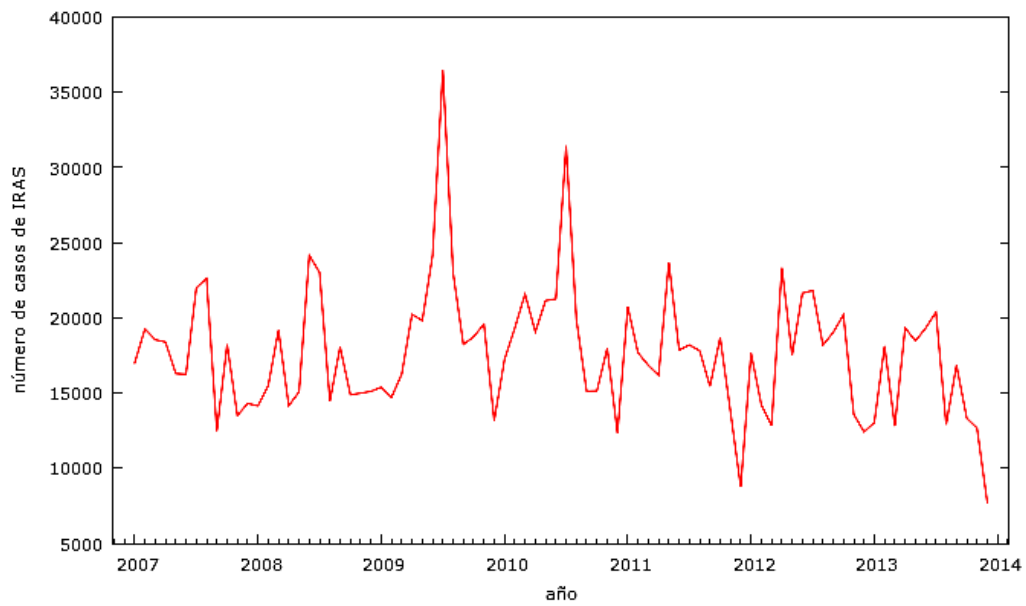
completo de cada una de las series, comienza en Enero del año 2007 hasta Diciembre del año 2013, obteniendo 84 observaciones en total por cada una de las enfermedades.

2.2. Obtención de los datos

La forma en como se han obtenidos los datos, ha sido mediante los registros de las enfermedades antes mencionadas, que las unidades de salud distribuidas en los 13 municipios de Santa Ana han aportado, así como también los registros proporcionados por los hospitales públicos, el seguro social y algunos hospitales privados con los cuales se han desarrollado acuerdos que permiten compartir este tipo de datos. Para comenzar con el estudio se presentará el gráfico de serie temporal de correspondiente a cada una de las enfermedades.

Gráfico de la serie de tiempo IRAS mensual desde Enero 2007 hasta Diciembre de 2013.

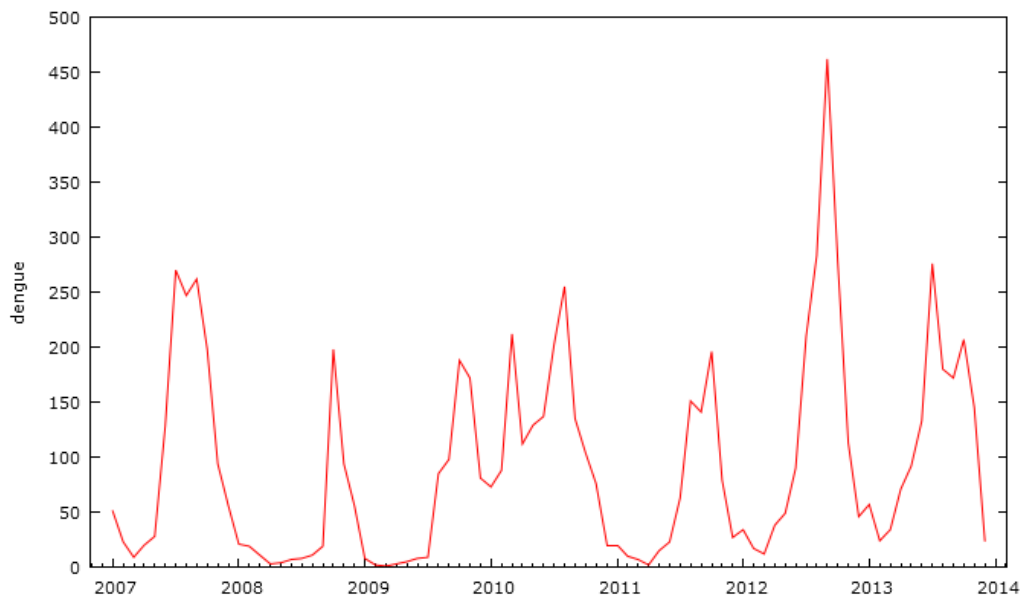
Figura 2.1: Gráfico de la serie de tiempo IRAS



fuelle: SIBASI Santa Ana

Gráfico de la serie de tiempo Dengue mensual desde Enero 2007 hasta Diciembre de 2013.

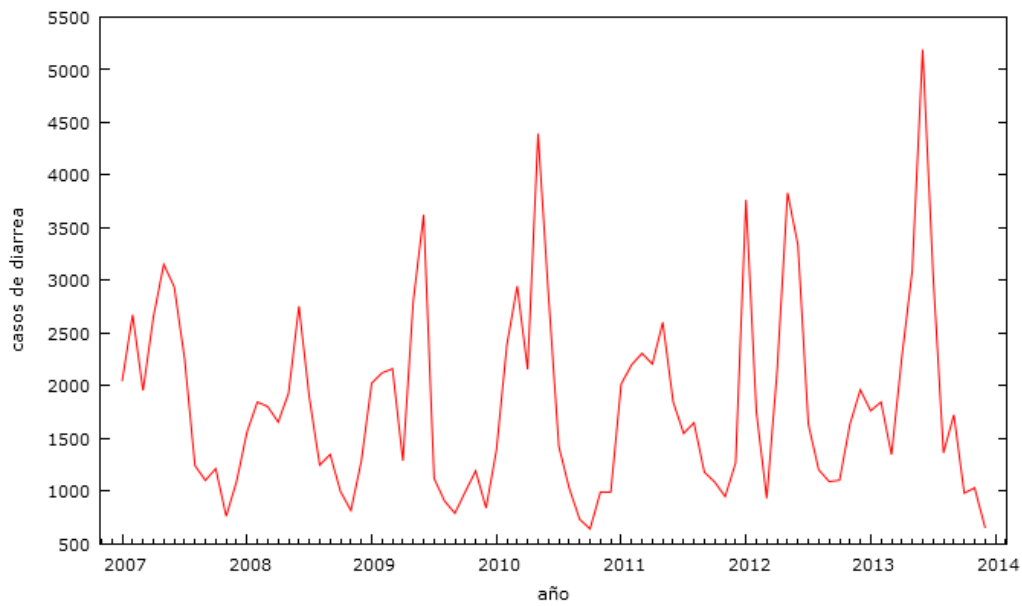
Figura 2.2: Gráfico de la serie de tiempo Dengue



fuelle: SIBASI Santa Ana

Gráfico de la serie de tiempo Diarrea mensual desde Enero 2007 hasta Diciembre de 2013.

Figura 2.3: Gráfico de la serie Diarrea



fuelle: SIBASI Santa Ana

3.1. Conceptos Básicos

Definimos una serie de tiempo como un conjunto de observaciones repetidas de la misma variable tal como: Y_1, Y_2, \dots, Y_T . A partir de esta caracterización se desprenden dos conceptos fundamentales en el análisis de series de tiempo. Los mismos se definen mediante la imposición de algunos supuestos sobre el comportamiento de Y_t .

1. Ruido blanco: el proceso t se define como ruido blanco si cumple con las siguientes condiciones:
 - a) La esperanza de Y_t o promedio es igual a 0 para todos los periodos t . Esto es $E(Y_t) = 0$ para toda t .
 - b) La varianza de Y_t es constante y por consiguiente independiente del tiempo. Esto es $var(Y_t) = \sigma^2$.
2. Independiente e idénticamente distribuido⁸ (iid): el proceso Y_t se dice ser iid si cumple con las siguientes condiciones:
 - a) La esperanza de Y_t es constante pero no necesariamente igual a 0 para todos los periodos t . Esto es $E(Y_t) = \mu$.
 - b) La varianza de Y_t es constante y por consiguiente independiente del tiempo. Esto es $var(Y_t) = \sigma^2$.

⁸iid: Independiente e idénticamente distribuido

c) Y_t es independiente de k y para todas las t y k con $t \neq k$.

Uno de los pasos más importantes en la selección de un método para pronosticar adecuado con datos de una serie de tiempo es considerar los diferentes tipos de patrones de datos. Existen cuatro tipos generales **estacionariedad, tendencia, estacionalidad y cíclicidad**.

3.1.1. Estacionariedad

Cuando los datos recopilados en el transcurso del tiempo fluctúan alrededor de un nivel o una media constante, hay un patrón horizontal o estacionario. Se dice que este tipo de series es estacionario en su media. En caso de que la serie no sea estacionaria en media, se hacen diferenciaciones (dato actual menos dato anterior) para lograr estabilizar la serie de tiempo. Y cuando la serie de tiempo no posee una variabilidad fija, es decir, una varianza constante se dice que la serie no es estacionaria en varianza, y para lograr identificar estacionariedad en varianza hay que utilizar el gráfico de rango-media que en el eje X lleva las medias por cada año (si la serie es mensual) y los rangos en el eje Y (también para cada año si la serie es mensual). Si los puntos del plano cartesiano se ajustan a una recta con pendiente positiva se dice que no hay estacionariedad en varianza, y para lograr la estacionariedad en varianza se utilizan las transformaciones de Box-Cox, en este estudio se utiliza la transformación logarítmica para estabilizar la varianza.

3.1.2. Tendencia

La **tendencia** es el componente de largo plazo que representa el crecimiento o el descenso en la serie de tiempo, durante un periodo extenso.

3.1.3. Estacionalidad

Cuando las observaciones se ven influidas por factores temporales, existe un patrón **estacional**. El componente **estacional** se refiere a un patrón de

cambio que se repite año tras año. Para las series mensuales, el componente estacional mide la variabilidad de la serie cada enero, cada febrero, y así sucesivamente. Para una serie trimestral, hay cuatro elementos estacionales, uno por cada trimestre.

3.1.4. Función de Autocorrelación (FAS ó FAC)

Cuando se mide una variable a lo largo del tiempo, las observaciones en diferentes periodos a menudo están relacionadas o correlacionadas. Esta correlación se mide usando el coeficiente de autocorrelación. **Autocorrelación** es la correlación que existe entre una variable retrasada uno o más periodos consigo misma.

Los patrones de datos que incluyen componentes como tendencia y estacionalidad pueden estudiarse usando autocorrelaciones. Los patrones se identifican examinando los coeficientes de autocorrelación de una variable en diferentes retrasos de tiempo.

La ecuación 3.1 es la fórmula para calcular el coeficiente de autocorrelación ρ_k entre las observaciones Y_t y Y_{t-k} , que se encuentran a k periodos de distancia.

$$\rho_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2}, k = 0, 1, 2, \dots \quad (3.1)$$

ρ_k : Es el coeficiente de autocorrelación para un retraso de k periodos.

\bar{Y} : Es la media de los valores de la serie de tiempo.

Y_t : Es la observación en el periodo t .

Y_{t-k} : Es la observación k periodos anteriores o durante un periodo $t - k$.

3.1.5. Función de Autocorrelación Parcial (PACF ó FAP)

La autocorrelación parcial mide la correlación entre dos variables separadas por k periodos cuando no se considera la dependencia creada por los

retardos intermedios existentes entre ambas.

$$\rho_{kk} = \text{corr}(\mathbf{Y}_t \mathbf{Y}_{t-k} | \mathbf{Y}_{t-1} \mathbf{Y}_{t-2} \mathbf{Y}_{t-3}, \dots, \mathbf{Y}_{t-k-1})$$

$$\rho_{kk} = \frac{\text{cov}(\mathbf{Y}_t - \widehat{\mathbf{Y}}_t, \mathbf{Y}_{t-k} - \widehat{\mathbf{Y}}_{t-k})}{\sqrt{\text{var}(\mathbf{Y}_t - \widehat{\mathbf{Y}}_t)} \sqrt{\text{var}(\mathbf{Y}_{t-k} - \widehat{\mathbf{Y}}_{t-k})}}$$

ρ_{kk} : Es el coeficiente de autocorrelación parcial.

$\widehat{\mathbf{Y}}_{t-k}$: valor estimado en el periodo $t - k$

\mathbf{Y}_t : Es la observación en el periodo t .

Los coeficientes de autocorrelación para diferentes retrasos de tiempo de una variable pueden usarse para contestar las siguientes preguntas acerca de una serie de tiempo:

1. Los datos son ruido blanco.
2. Los datos muestran una tendencia (son no estacionarios).
3. Los datos son estacionarios.
4. Los datos son estacionales.

Si una serie es **aleatoria**, las autocorrelaciones entre \mathbf{Y}_t y \mathbf{Y}_{t-k} para cualquier retraso de tiempo k son cercanas a cero. Los valores sucesivos de una serie de tiempo no están relacionados entre sí. Si una serie muestra una tendencia, las observaciones sucesivas están altamente correlacionadas y es típico que los coeficientes de correlación sean significativamente diferentes de cero, para los primeros retrasos de tiempo, y de forma gradual tienden a cero conforme se incrementa el número de retrasos.

Si una serie tiene un **patrón estacional**, se presentará un coeficiente de autocorrelación significativo en el retraso de tiempo estacional o en los múltiplos

del retraso estacional. El retraso estacional se considera de 4 para datos trimestrales y 12 para datos mensuales.

¿ Cómo determina un analista si un coeficiente de autocorrelación es significativamente diferente de cero ?

Quenouille (1949) y otros han demostrado que los coeficientes de autocorrelación de datos aleatorios tienen una distribución muestral que puede aproximarse mediante una curva normal, con una media de cero y una desviación estándar aproximada de $\frac{1}{\sqrt{n}}$. Sabiendo esto, el analista compara los coeficientes de autocorrelación de la muestra, con esta distribución teórica de la muestra y determinar si, para retrasos de tiempo dados, provienen de una población cuya media sea cero. En realidad, algunos paquetes de software usan una fórmula ligeramente diferente, como se indica en la ecuación 4.5, para calcular las desviaciones estándar (o errores estándar) de los coeficientes de autocorrelación. Esta fórmula supone que cualquier autocorrelación antes del retraso de tiempo k es diferente de cero, y que cualquier autocorrelación anterior al retraso k es cero. Para una autocorrelación en el retraso 1, se utiliza el error estándar

$$SE(r_k) = \sqrt{\frac{1 + 2 \sum_{i=1}^{k-1} r_i^2}{n}} \quad (3.2)$$

donde $SE(r_k)$: error estándar (desviación estándar estimada) de la autocorrelación en retraso k

r_i : la autocorrelación en el retraso i .

k : es el retraso de tiempo.

n : número de observaciones en la serie de tiempo.

Si la serie es verdaderamente **aleatoria**, casi todos los coeficientes de autocorrelación de la muestra deberían estar dentro de un rango especificado por cero, más o menos cierto número de errores estándar. En un nivel de con-

fianza específico, una serie puede considerarse aleatoria, si cada uno de los coeficientes de autocorrelación calculados se encuentra dentro del intervalo alrededor de 0, definido por $\mathbf{0} \pm tSE(r_k)$, donde el multiplicador t es un punto porcentual adecuado de una distribución t . Si bien es útil cada prueba de r_k para saber si, en lo individual, es significativamente diferente de 0, también es recomendable examinar un conjunto de r_k consecutivos, como un grupo. Podemos usar una prueba de conjunto para ver si, por ejemplo, de los primeros 10 valores de r_k son significativamente diferentes de un conjunto donde los 10 valores son cero. Una prueba de baúl común es el estadístico modificado Q de Ljung-Box 3.3. Esta prueba se aplica usualmente a los residuos de un modelo de pronóstico. Si las autocorrelaciones se calculan mediante un proceso aleatorio (o de ruido blanco), el estadístico Q tiene una distribución chi cuadrada con m grados de libertad (el número de retrasos de tiempo por probar). Sin embargo, para los residuos de un modelo de pronóstico, el estadístico Q tiene una distribución chi cuadrada, con m grados de libertad menos el número de parámetros estimados en el modelo. La hipótesis nula se refiere a que todas las autocorrelaciones en el conjunto son cero. Alternativamente, el valor p generado por el estadístico de prueba Q puede ser calculado e interpretado. El estadístico Q está dado por la ecuación 3.3

$$Q = n(n + 2) \sum_{k=1}^m \frac{r_k^2}{n - k} \quad (3.3)$$

donde

n : número de observaciones en la serie de tiempo

k : retraso de tiempo

m : número de retrasos de tiempo que se van a probar

r_k : la función de autocorrelación muestral de los residuos atrasados k periodos

¿Cómo identificar cuando los datos de una serie son aleatorios?

Un modelo aleatorio simple, a menudo llamado modelo de ruido blanco, está representado por la ecuación 3.4. La observación \mathbf{Y}_t se compone de dos partes: \mathbf{c} , el nivel general; y $\boldsymbol{\varepsilon}_t$, el componente de error aleatorio⁹. Es importante advertir que se supone que el componente $\boldsymbol{\varepsilon}_t$ no está correlacionado de un periodo a otro.

$$\mathbf{Y}_t = \mathbf{c} + \boldsymbol{\varepsilon}_t \quad (3.4)$$

¿Cómo identificar cuando los datos de una serie muestran tendencia?

Si una serie muestra una tendencia, hay una relación significativa entre los valores sucesivos de la serie de tiempo. Los coeficientes de autocorrelación son usualmente grandes para varios de los primeros retrasos de tiempo y luego, conforme se incrementa el número de retrasos, caen gradualmente hacia cero. Una serie de tiempo estacionaria es aquella cuyas propiedades estadísticas básicas, como la media y la varianza, permanecen constantes en el tiempo. Por lo tanto, se dice que una serie que varía alrededor de un nivel fijo (sin crecimiento ni decrecimiento) con el paso del tiempo es estacionaria. Se dice también que una serie que contiene una tendencia es no estacionaria. Los coeficientes de autocorrelación de una serie estacionaria decrecen hacia cero bastante rápidamente, por lo común después del segundo o tercer retraso de tiempo. Por otro lado, las autocorrelaciones muestrales de serie no estacionaria se permanecen muy grandes durante varios periodos. A menudo, para analizar la serie no estacionaria, se elimina la tendencia antes de aplicar cualquier modelo. Con frecuencia se utiliza un método llamado **diferencias**¹⁰

⁹El error aleatorio es la diferencia entre un dato estimado y su correspondiente dato real

¹⁰El número de diferenciaciones se denota por d , $\mathbf{d} = \mathbf{Y}_t - \mathbf{Y}_{t-1}$ raras veces se requiere de más de una diferenciación en una serie tiempo

para eliminar la tendencia de una serie no estacionaria este metodo consiste en restarle al dato actual de la serie el dato anterior, lo cual permite estabilizar la serie de tiempo.

¿Cómo identificar estacionalidad en una serie temporal?

Si una serie es estacional, un patrón relacionado con el calendario se repite así mismo durante un intervalo de tiempo específico (generalmente un año). Las observaciones de la misma posición, en diferentes periodos estacionales, tienden a estar relacionadas. Si se analizan datos trimestrales que tienen un patrón estacional, los primeros trimestres tienden a parecerse, los segundos trimestres tienden a parecerse, y así sucesivamente, y habrá un coeficiente de autocorrelación significativo en el retraso de tiempo 4. Si se analizan datos mensuales, aparecerá un coeficiente de autocorrelación significativo en el retraso de tiempo 12. Es decir, enero se correlacionará con otros eneros, febrero se correlacionará con otros febreros y así sucesivamente.

3.2. Principales modelos ARIMA de procesos de series temporales

3.2.1. Introducción

Este capítulo presenta una clase de modelos que generan pronósticos exactos con base en una descripción de patrones históricos de los datos. Los modelos autorregresivos integrados de promedio móvil (ARIMA) son una clase de modelos lineales que son capaces de representar tanto series de tiempo estacionarias como no estacionarias. Recuerde que los procesos estacionarios varían alrededor de un nivel fijo, en tanto que los procesos no estacionarios carecen de un nivel promedio constante natural. Los modelos ARIMA no implican variables independientes en su construcción. En vez de ello, utilizan la información de la serie misma para generar los pronósticos. Por ejemplo,

un modelo ARIMA para ventas mensuales proyectaría el patrón de ventas históricas para generar un pronóstico para las ventas del siguiente mes.

Los modelos ARIMA dependen mucho de los patrones de autocorrelación en los datos. La metodología para identificar, ajustar y verificar los modelos ARIMA adecuados fue impulsada fuertemente por dos especialistas en estadística, G. E. P. Box y G. M. Jenkins. Por esta razón, los modelos ARIMA para la generación de pronósticos se conocen a menudo como metodología Box-Jenkins.

3.2.2. Metodología Box-Jenkins

La metodología Box-Jenkins para generar pronósticos es diferente de la mayoría de los métodos porque no supone ningún patrón particular en los datos históricos de las series que se van a pronosticar. Se basa en un enfoque iterativo para identificar un modelo posible a partir de una clase general de modelos. Luego, el modelo seleccionado se coteja con los datos históricos para ver si describe la serie con exactitud. El modelo está bien ajustado si los residuos son generalmente pequeños, están distribuidos aleatoriamente y no contienen información útil. Si el modelo especificado no es satisfactorio, el proceso se repite usando un nuevo modelo diseñado para mejorar el original. Este procedimiento iterativo continúa hasta que se encuentra un modelo satisfactorio. En ese momento, el modelo se considera útil para pronosticar. La selección inicial de un modelo ARIMA se basa en examinar una gráfica de la serie de tiempo (para observar su carácter general) y en analizar sus autocorrelaciones para varios retrasos de tiempo. Específicamente, el patrón de las autocorrelaciones muestrales calculado a partir de la serie de tiempo se coteja con el patrón conocido de autocorrelación asociado con un modelo ARIMA particular. Este acoplamiento se hace tanto para las autocorrelaciones como para las autocorrelaciones parciales. Una autocorrelación parcial

en el retraso de tiempo k es la correlación entre \mathbf{Y}_t y \mathbf{Y}_{t-k} (las respuestas para los periodos t y $t - k$, respectivamente) después de ajustar los efectos de los valores que intervienen, $\mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \dots, \mathbf{Y}_{t-k}$. Newbold y Bos (1994), así como Box, Jenkins y Reinsel (1994) se encargaron de derivar y estudiar las autocorrelaciones parciales.

3.2.3. Modelos Autorregresivos(AR)

Un modelo autorregresivo de orden p tiene la forma

$$\mathbf{Y}_t = \phi_0 + \phi_1 \mathbf{Y}_{t-1} + \phi_2 \mathbf{Y}_{t-2} + \dots + \phi_p \mathbf{Y}_{t-p} + \varepsilon \quad (3.5)$$

donde \mathbf{Y}_t : es la variable respuesta(o dependiente) en el tiempo t

$\mathbf{Y}_{t-1}, \dots, \mathbf{Y}_{t-p}$: variable de respuesta en los retrasos de tiempo $t - 1, t - 2, \dots, t - p$, respectivamente; estas \mathbf{Y} desempeñan el papel de variables independientes.

$\phi_0, \phi_1, \phi_2, \dots, \phi_p$: coeficientes que serán estimados

¹¹ ε : término de error en el tiempo t , el cual representa los efectos de variables no explicadas por el modelo; los supuestos acerca del término de error son los mismos que los del modelo de regresión estándar (media 0 y varianza constante). El modelo de la ecuación 3.5 tiene la apariencia de un modelo de regresión con valores retrasados de la variable dependiente en las posiciones de la variable independiente, de aquí el nombre de modelo autorregresivo. Los modelos autorregresivos son modelos adecuados para series de tiempo estacionarias, y el coeficiente ϕ_0 está relacionado con el nivel constante de la serie. Si los datos varían alrededor de cero o se expresan como desviaciones de la media, no se requiere el coeficiente ϕ_0 .

Para modelos autorregresivos, los pronósticos dependen de los valores observados en periodos anteriores. Para modelos AR(2), los pronósticos del

¹¹El coeficiente ϕ_0 está relacionado con la media del proceso, μ , mediante

$$\phi_0 = \mu(1 - \phi_1 - \phi_2 - \phi_3 - \dots - \phi_p)$$

siguiente valor dependen de los valores observados en los dos periodos anteriores. Para modelos AR(3), los pronósticos del siguiente valor dependen de los valores observados en los tres periodos anteriores, etcétera.

3.2.4. Modelos de Medias Móviles(MA)

Un modelo de promedio móvil de q-ésimo orden tiene la forma

$$Y_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon \quad (3.6)$$

donde Y_t : es la variable respuesta(o dependiente) en el tiempo t

$\varepsilon_{t-1}, \dots, \varepsilon_{t-q}$: Errores en periodos anteriores que, para el tiempo t , se incorporan a la respuesta, Y_t

μ : Media constante del proceso.

$\theta_1, \theta_2, \dots, \theta_q$: Coeficientes que serán estimados

ε_t : Término de error, el cual representa los efectos de las variables no explicadas por el modelo; los supuestos acerca del término de error son los mismos que aquellos del modelo de regresión estándar.

La ecuación 3.6 es similar a la ecuación 3.5, excepto que la variable dependiente, Y_t , depende de los valores anteriores de los errores y no de los valores de la variable misma. Los modelos de promedio móvil (MA) permiten hacer pronósticos de Y_t con base en una combinación lineal de un número finito de errores pasados, mientras que los modelos autorregresivos (AR) pronostican Y_t como una función lineal de un número finito de valores pasados de Y_t . El término promedio móvil para el modelo de la ecuación 3.6 es histórico. Aquí, promedio móvil se refiere al hecho de que la desviación de la respuesta de su media, $Y_t - \mu$, es una combinación lineal de los errores actuales y pasados y que, conforme el tiempo avanza, los errores implicados en esta combinación lineal también se mueven hacia delante.

Un modelo de promedio móvil de q-ésimo orden tiene la forma

$$\begin{aligned}
 Y_t - \mu &= \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon \\
 Y_{t+1} - \mu &= \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q+1} + \varepsilon
 \end{aligned}$$

Los pesos $\theta_1, \theta_2, \dots, \theta_q$ no necesariamente suman 1 y pueden ser positivos o negativos, aun cuando estén precedidos de un signo negativo en la especificación del modelo, los términos de error se pueden agregar en secuencia para obtener un $MA(q)$, donde q es el número de términos. Los coeficientes de autocorrelación para el modelo $MA(1)$ caen a cero después del primer retraso, mientras que los coeficientes de autocorrelación parcial tienden a cero gradualmente. Además, los coeficientes de autocorrelación del modelo $MA(2)$ se vuelven cero después del segundo retraso de tiempo, mientras que las autocorrelaciones parciales tienden gradualmente a cero. Otra vez, debemos mencionar que las funciones de autocorrelación muestrales van a diferir de estas funciones teóricas a causa de la variación de la muestra.

3.2.5. Modelos de promedio móvil autorregresivos

Un modelo con términos autorregresivos se puede combinar con un modelo que tenga términos de promedio móvil para obtener un modelo mixto de promedio móvil autorregresivo. Para representar estos modelos, es conveniente utilizar la notación $ARMA(p, q)$, donde p es el orden de la parte autorregresiva y q es el orden de la parte del promedio móvil. Un modelo $ARMA(p,q)$ tiene la forma general

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon \quad (3.7)$$

Los modelos **ARMA** (p,q) pueden describir una gran variedad de comportamientos de las series de tiempo estacionarias.¹² Los pronósticos generados por un modelo **ARMA** (p,q) dependen de los valores actuales y pasados de la respuesta Y , así como de los valores actuales y pasados de los errores (residuos), e .

3.3. Modelos Autoregresivo Integrados de Medias Móviles **ARMA** (p,d,q)

Los modelos de series de tiempo analizados hasta ahora se basan en el supuesto de estacionariedad, esto es, la media y la varianza para una serie de tiempo son constantes en el tiempo y la covarianza es invariante en el tiempo. Pero se sabe que muchas series de tiempo y en especial las series económicas no son estacionarias, porque pueden ir cambiando de nivel en el tiempo o sencillamente la varianza no es constante en el tiempo, a este tipo de proceso se les considera procesos integrados. Por consiguiente, se debe diferenciar una serie de tiempo d veces para hacerla estacionaria y luego aplicarla a esta serie diferenciada un modelo **ARMA** (p, q) , se dice que la serie original es **ARIMA** (p, d, q) , es decir, una serie de tiempo autoregresiva integrada de media móvil. Donde p denota el número de términos autoregresivos, d es el número de veces que la serie debe ser diferenciada para hacerla estacionaria y q el número de términos de la media móvil.

Su expresión algebraica es:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d Y_t = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \epsilon_t$$

$B^k(Y_t) = Y_{t-k}$ es el operador de retardos o retroactivo.

¹²Observe que cuando $q = 0$, el modelo **ARMA** $(p, 0)$ se reduce a un modelo autorregresivo puro de orden p . De forma similar, cuando $p = 0$, el modelo **ARMA** $(0, q)$ es un modelo de promedio móvil puro de orden q .

3.4. Proceso Estacional Autoregresivo Integrado y de Media Móvil

Cuando una serie de tiempo en estudio tiene intervalos de observación menores a un año, entonces es frecuente que estas tengan variaciones ó patrones sistemáticos cada cierto periodo, estas variaciones sistemáticas inferiores a un año por ejemplo semestral, mensual, diario, etc. Deben ser captadas en los llamados Factores Estacionales, dentro de la estructura del modelo a construirse. cada una de estas series puede ser estacionaria o no estacionaria. de esta manera se combinan términos ordinarios del proceso ARMA y términos estacionales, así como diferencias regulares y diferencias estacionales para transformar en series estacionarias, esto es $\Delta_s^D \Delta^d Y_t$. Este tipo de procesos tiene las siguientes características:

- 1 Contiene una componente $ARIMA(p, d, q)$ que modela la dependencia regular, que es la dependencia asociada a observaciones consecutivas.
- 2 Contiene una componente $ARIMA(P, D, Q)$ que modela la dependencia estacional, que está asociada a observaciones separadas por periodos.

La estructura general de un modelo $ARIMA(p, d, q)(P, D, Q)_s$, es:

$$\begin{aligned}
 Y_t = & \underbrace{c + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p}}_{AR(p)} + \underbrace{\Phi_1 Y_{t-s} + \Phi_2 Y_{t-2s} + \dots + \Phi_P Y_{t-Ps}}_{SAR(P)} \\
 & + \underbrace{\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}}_{MA(q)} - \underbrace{\Theta_1 \varepsilon_{t-s} - \dots - \Theta_Q \varepsilon_{t-Qs}}_{SMA(Q)}
 \end{aligned}$$

$SAR(P)$: Es el proceso estacional autorregresivo de orden P .

$SMA(Q)$: Es el proceso estacional de medias móviles de orden Q .

Los parámetros son :

$$\phi_1, \phi_2, \dots, \phi_p, \Phi_1, \Phi_2, \dots, \Phi_P, \theta_1, \theta_2, \dots, \theta_q, \Theta_1, \Theta_2, \dots, \Theta_q, \quad \varepsilon_t \sim N(0, \sigma^2)$$

3.5. Criterios de Información

Es posible identificar (seleccionar) los modelos ARIMA examinando una gráfica de la serie y al establecer una correspondencia entre los patrones de las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra con los patrones teóricos conocidos de los procesos ARIMA. Sin embargo, existe cierta subjetividad implicada en este procedimiento, y es posible que dos (o más) modelos iniciales sean congruentes con los patrones de las autocorrelaciones de la muestra y las autocorrelaciones parciales de la muestra. Más aún, después de la estimación y verificación, ambos modelos pueden representar adecuadamente los datos. Si los modelos contienen el mismo número de parámetros, el modelo con menor error cuadrático medio, es por lo general el preferido. Si los modelos contienen diferentes números de parámetros, el principio de parsimonia conduce a la selección del modelo más sencillo. Sin embargo, el modelo con más parámetros puede tener un error cuadrático medio considerablemente menor. Se ha desarrollado un método para la selección del modelo que considere tanto el ajuste del modelo como el número de parámetros. El criterio de **información de Akaike**¹³ (1974), o AIC, selecciona el mejor modelo de un grupo de modelos candidatos que minimiza

$$AIC = \ln(\hat{\sigma}^2) + \frac{2}{n}r \quad (3.8)$$

donde:

\ln : logaritmo natural

¹³Akaike Information Criterion

$\hat{\sigma}^2$: suma residual de cuadrados dividida entre el número de observaciones

n : número de observaciones (residuos)

r : número total de parámetros (incluyendo el término constante) en el modelo

ARIMA

El criterio de **información bayesiano** que desarrolló Schwarz (1978), o BIC¹⁴, selecciona el modelo que minimiza

$$BIC = \ln(\hat{\sigma}^2) + \frac{\ln(n)}{n}r \quad (3.9)$$

Ambos estadísticos miden la bondad del ajuste del modelo. Es decir, los modelos más adecuados son los que tienen mejor ajuste. Para ello, es suficiente verificar que el modelo que mejor ajusta es el que presenta el AIC o BIC más bajo. En términos generales, el criterio AIC es más propio de los modelos Autorregresivos, mientras que el BIC es un criterio mucho más general.

3.6. Estrategia de implementación de la construcción del modelo

El método Box-Jenkins se basa en una estrategia iterativa para la construcción del modelo que consiste en la selección de un modelo inicial (identificación del modelo), la estimación de los coeficientes del modelo (estimación de parámetros) y el análisis de los residuos (verificación del modelo). Si es necesario, el modelo inicial se modifica y el proceso se repite hasta que los residuos indiquen que ya no es necesaria otra modificación. En este punto, estamos en condiciones de utilizar el modelo ajustado para pronosticar. A continuación se estudian con detalle los pasos de la estrategia para la construcción del modelo.

¹⁴Bayesian Information Criterion

3.6.1. Paso 1: Identificación del modelo

- 1 El primer paso en la identificación del modelo es determinar si la serie es estacionaria, es decir, si la serie de tiempo parece variar alrededor de un nivel fijo. Es útil observar una gráfica de la serie junto con la función de autocorrelación de la muestra. Para identificar una serie no estacionaria se observa si el gráfico de la serie parece crecer o declinar en el tiempo y las autocorrelaciones de la muestra no se desvanecen rápidamente. Si la serie es no estacionaria, a menudo se puede convertir en una serie estacionaria por diferenciación. Es decir, la serie original se sustituye por una serie de las diferencias (al dato actual se le resta el anterior).

La diferenciación se hace hasta que una gráfica de los datos indica que la serie varía alrededor de un nivel fijo y las autocorrelaciones de la muestra desaparecen con rapidez. El número de diferencias requeridas para alcanzar la estacionariedad se denota por d .

Los modelos de las series no estacionarias se llaman modelos de promedio móvil integrados autorregresivos y se denotan por **ARIMA(p,d,q)**. Aquí, p indica el orden de la parte autorregresiva, d indica el número de diferenciaciones, y q indica el orden de la parte de promedio móvil. Si la serie original es estacionaria, entonces $d = 0$, y los modelos ARIMA se reducen a modelos ARMA¹⁵. Por consiguiente, de ahora en adelante, la notación ARIMA (p, d, q)¹⁶ se usará para identificar modelos tanto de series estacionarias ($d = 0$) como no estacionarias ($d > 0$).

¹⁵Cuando se usa un modelo ARMA para una serie diferenciada, tal vez no se requiera el término constante, ϕ_0 .

¹⁶El término integrado(d) significa que las diferencias se deben sumar (o integrar) para obtener la serie original.

- 2 Una vez que se ha obtenido una serie estacionaria, el analista debe identificar la forma del modelo que se utilizará. Cada modelo ARIMA tiene un conjunto único de autocorrelaciones y autocorrelaciones parciales, y el analista debe tener la capacidad de identificar los valores que se obtienen a partir de la muestra a uno de los patrones teóricos. Es posible que exista alguna ambigüedad en la identificación de un modelo ARIMA adecuado a partir de los patrones de las autocorrelaciones y las autocorrelaciones parciales. Por ello, la selección del modelo inicial deberá considerarse como tentativa. Se deben realizar análisis durante los pasos 2 y 3 para determinar si el modelo es adecuado. Si no lo es, la identificación de la forma del modelo se lleva a cabo comparando las autocorrelaciones y las autocorrelaciones parciales calculadas con los datos de las autocorrelaciones y las autocorrelaciones parciales teóricas de los diferentes modelos ARIMA.

Hay que tomar en cuenta que, si las autocorrelaciones de la muestra se desvanecen exponencialmente hacia cero y las autocorrelaciones parciales de la muestra se cortan de forma abrupta, el modelo requerirá términos autorregresivos. Si las autocorrelaciones de la muestra se cortan manera abrupta y las autocorrelaciones parciales de la muestra se desvanecen, el modelo requerirá términos de promedio móvil. Si tanto las autocorrelaciones de la muestra como las autocorrelaciones parciales de la muestra se desvanecen, tanto los términos autorregresivos como los términos de promedio móvil están indicados. Al contar el número de autocorrelaciones y autocorrelaciones parciales significativas de la muestra, pueden determinarse los órdenes de las partes MA y AR. Para juzgar su significancia, tanto las autocorrelaciones de la muestra como las autocorrelaciones parciales de la muestra se comparan gene-

ralmente con $\pm \frac{2}{\sqrt{n}}$ donde n es el número de observaciones en la serie de tiempo. Estos límites funcionan bien cuando n es grande.

En iguales condiciones, son preferibles los modelos simples en vez de los modelos complejos. Esto se conoce como el **principio de parsimonia**. Con una cantidad limitada de datos es relativamente fácil encontrar un modelo con un número grande de parámetros que se ajusten bien a los datos. Sin embargo, los pronósticos de un modelo probablemente van a ser deficientes porque buena parte de la variación en los datos se debe al error aleatorio que se está modelando. El objetivo es desarrollar el modelo más sencillo que ofrezca una descripción adecuada de las principales características de los datos.

El **principio de parsimonia** se refiere a la preferencia por los modelos sencillos por encima de los modelos complejos.

3.6.2. Paso 2: Estimación del modelo

- 1 Una vez que se ha seleccionado un modelo tentativo, se deben estimar los parámetros para ese modelo. Los parámetros de los modelos ARIMA se estiman minimizando la suma de cuadrados de los errores de ajuste. Estos mínimos cuadrados estimados deben obtenerse, en general, usando un procedimiento no lineal de mínimos cuadrados. Un procedimiento no lineal de mínimos cuadrados es simplemente un algoritmo que obtiene el mínimo de la función de la suma de los errores al cuadrado. Una vez que se determinan las estimaciones de los mínimos cuadrados y sus errores estándar, se pueden construir e interpretar los valores t del modo habitual. Los parámetros que son considerados significativamente diferentes de cero se conservan en el modelo ajustado; los parámetros que no son significativos se eliminan del modelo. Si los

parámetros tienen un estadístico t menor que 2 se eliminan del modelo.

- 2 Se calcula el error cuadrático medio de los residuos, una estimación de la varianza del error.

El error cuadrático medio de los residuos se define como

$$s^2 = \frac{\sum_{i=1}^n e_t^2}{n - r}$$

$$s^2 = \frac{\sum_{i=1}^n (Y_t - \widehat{Y}_t)^2}{n - r}$$

donde

e_t : Resíduo para el tiempo t

n : número de resíduos

r : número total de parámetros estimados

El error cuadrático medio de los residuos es útil para evaluar el ajuste y comparar modelos diferentes. También es útil para calcular los límites de error de pronóstico.

3.6.3. Paso 3: Verificación del modelo(Diagnosis)

Esta etapa también es llamado Chequeo o Diagnosis, antes de utilizar el modelo para pronosticar, debe verificarse qué tan adecuado es. En esencia, un modelo es adecuado si los residuos no se pueden usar para mejorar los pronósticos. Es decir, los residuos deben ser aleatorios.

- 1 Muchas de las gráficas residuales que son útiles en el análisis de regresión se pueden desarrollar para los residuos de un modelo ARIMA. Un histograma y una gráfica de probabilidad normal (para verificar la

normalidad), así como una gráfica de la secuencia del tiempo (para verificar los valores atípicos) son particularmente útiles. Se debe verificar que la Gráfica secuencial de los residuos sea estacionaria, también los residuos debe tener media 0 y varianza constante (ser ruido blanco).

- 2 Las autocorrelaciones residuales individuales deben ser pequeñas y generalmente dentro de la banda que está dentro del intervalo $\pm \frac{2}{\sqrt{n}}$ y muy próximas a cero. Las autocorrelaciones residuales significativas en retrasos cortos o retrasos estacionales sugieren que el modelo es inadecuado y que debe seleccionarse un nuevo modelo o modificarse.
- 3 Las autocorrelaciones residuales como un grupo deben ser congruentes con aquellas producidas por los errores aleatorios. Una verificación general de la idoneidad del modelo se realiza mediante una prueba de distribución chi cuadrada (χ^2) con base en el estadístico Q de **Ljung-Box**. Esta prueba considera los tamaños de las autocorrelaciones residuales como un grupo. El estadístico de prueba Q es

$$Q = n(n + 2) \sum_{k=1}^m \frac{r_k^2(e)}{n - k} \quad (3.10)$$

el cual está distribuido aproximadamente como una variable aleatoria chi cuadrada con $m - r$ grados de libertad, donde r es el número total de parámetros estimados en el modelo ARIMA.

En la ecuación 3.10

$r_k^2(e)$: autocorrelación residual para el retraso k

n : número de residuos

k : retraso de tiempo

m : número de retrasos de tiempo que van a ser evaluados

Si el valor p asociado con el estadístico Q es pequeño (digamos, el valor $p < 0,05$), el modelo se considera inadecuado. El analista debe considerar

un nuevo modelo o uno modificado y continuar el análisis hasta encontrar un modelo satisfactorio.

3.6.4. Paso 4: Elaboración de pronósticos

- 1 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. En general, para un nivel de confianza dado, cuanto más largo sea el tiempo guía del pronóstico, mayor será el intervalo de predicción. Esto es razonable, puesto que se espera que la incertidumbre sea mayor para el pronóstico de un valor distante que para el pronóstico de, digamos, la siguiente observación. El cálculo de los pronósticos e intervalos de predicción es una labor tediosa y es preferible dejarla a la computadora. Los programas de computadora que ajustan modelos ARIMA generan pronósticos e intervalos de predicción a requerimiento del analista.
- 2 Conforme más datos están disponibles, se puede usar el mismo modelo ARIMA para generar pronósticos modificados de otro origen de tiempo. Es una buena idea hacer un seguimiento de los errores de pronóstico. Si los errores más recientes tienden a ser sistemáticamente más grandes que los errores anteriores, tal vez sea el momento de reevaluar el modelo. En este punto, es posible que se requiera otra estrategia de iteración para construir el modelo. Lo mismo es verdadero si los errores recientes del pronóstico tienden a ser positivos de manera sistemática (predicen una respuesta muy baja) o negativos (predicen una respuesta muy alta).

3.7. Metodología

3.7.1. Tipo de estudio

Se realiza un estudio descriptivo retrospectivo porque es un estudio longitudinal en el tiempo que se analiza en el presente, pero con datos del pasado.

3.7.2. Población

La población objeto de estudio involucrada en este trabajo está formada por todos los habitantes del Departamento de Santa Ana.

3.7.3. Muestra

La muestra consta del número de personas que adquirieron las enfermedades Dengue, Diarrea e IRAS por cada año desde 2007 hasta 2013.

3.7.4. Instrumentos

Con el propósito de obtener un modelo que se ajuste a las series de tiempo se utilizan softwares para determinar el mejor modelo, éstos son: **Gretl**, **R**, **SPSS** y **Eviews**.

3.7.5. Procedimiento

Para presentar las particularidades de la utilización de la Modelación ARIMA en la generación de pronósticos para tres enfermedades transmisibles en el Departamento de Santa Ana en el periodo comprendido entre los años 2007-2013, las enfermedades que se tratan son Dengue, Diarre e Infecciones Respiratorias Agudas(IRAS). Previo a la utilización de la metodología ARIMA se verifica que las series cumplieran con los requisitos recomendados por algunos autores: consistencia, estabilidad y no existencia de valores perdidos y que tuvieran un número suficiente de observaciones (más de 50 observaciones por cada serie temporal). La existencia de autocorrelación se explora mediante el gráfico de las funciones de autocorrelación o correlograma

y los órdenes de los términos incluidos en los diferentes modelos se correspondieron con los coeficientes de correlación simple y parcial estadísticamente significativos. Para la selección del modelo adecuado se sigue la metodología Box-Jenkins que consta de las siguientes etapas :Identificación, Estimación, Diagnósis o chequeo y Predicciones (Pronósticos), se sigue ésta metodología siempre que la serie sea estacionaria en media y en varianza. Para lograr la estacionariedad de la serie y estabilizar la media y las autocorrelaciones se realizan diferenciaciones, es decir reemplazar las observaciones por el valor de la diferencia entre ésta y la observación anterior para la diferenciación regular (de orden 1) o reemplazarla por la diferencia entre la observación actual y la correspondiente al periodo estacional anterior para la diferenciación estacional. Además se efectúa, sobre todo transformación logarítmica de los datos para la estabilización de la varianza. La búsqueda del modelo que mejor se ajuste a las series se efectúa tratando de que sean sencillos, explicaran bien las correlaciones internas del proceso y produjeran una diferencia mínima entre el valor observado y el esperado. Esto se realiza de manera iterativa, probando las combinaciones más lógicas de términos regulares AR y MA y de términos estacionales SAR y SMA, incluyendo las diferenciaciones necesarias. La adecuación de un modelo se estableció a partir del análisis de la significación estadística de los parámetros del mismo, de las características de los residuos o errores producidos por el modelo sobre la serie y del valor del error de predicción, se utiliza el Porcentaje de Error Medio Absoluto (MAPE).

Capítulo 4

Ajuste del Modelo

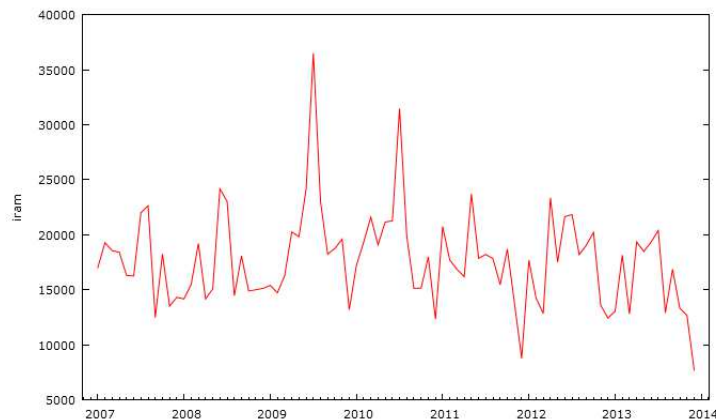
4.1. Identificación del modelo de la serie IRAS

En esta etapa se busca que la serie de tiempo sea estacionaria en media y en varianza para poder aplicar la Metodología Box-Jenkins.

4.1.1. Estacionariedad en media

Se inicia la etapa de identificación con la serie de tiempo IRAS, a continuación se muestra el gráfico de la serie temporal

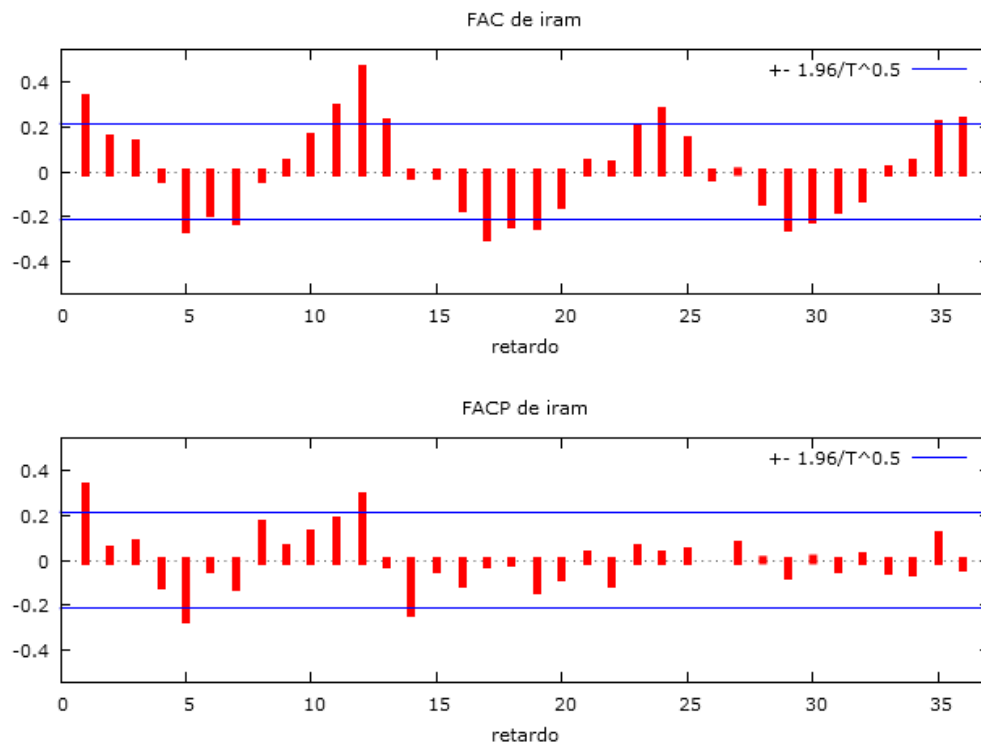
Figura 4.1: Gráfico de la serie de tiempo IRAS desde Enero 2007 hasta Diciembre 2013



En el gráfico 4.1 se puede observar que la serie oscila en torno a un nivel promedio más o menos constante (no posee tendencia o tramos con medias distintas). Se puede concluir, por lo tanto que la serie IRAS es ya estacionaria en media y no es preciso realizar transformaciones o diferenciación regular. Analizando los correlogramas de IRAS en el gráfico 4.2 se obtiene la misma

conclusión ya que el correlograma de IRAS decrece rápidamente hacia cero en forma senos y cosenos tanto para función de Autocorrelación(FAC) como para la función de Autocorrelación Parcial FACP

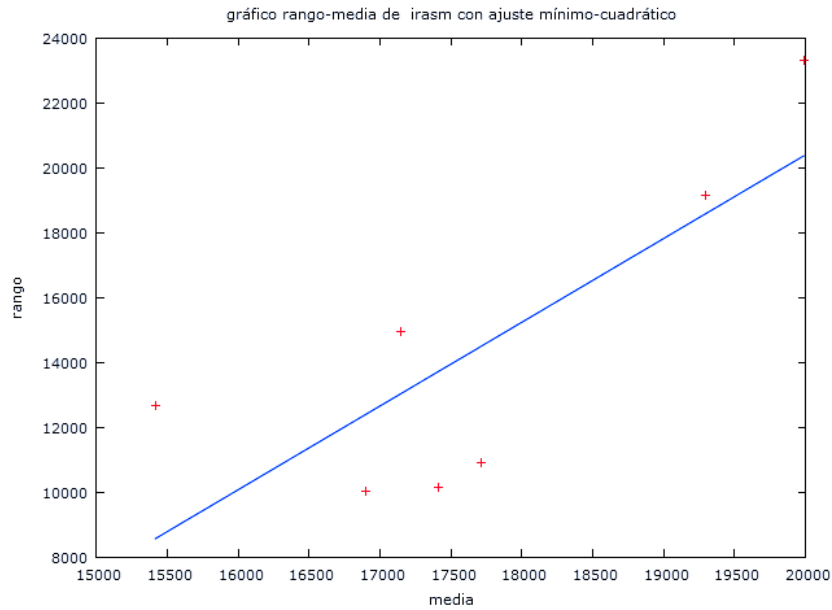
Figura 4.2: Correlograma de la serie IRAS sin transformaciones



Para contrastar la estacionariedad en varianza se observa que en el gráfico 4.1 la variabilidad crece entre los años 2009 y 2011, por lo tanto la serie IRAS no es estacionaria en varianza. Observando el gráfico de rango-media se puede confirmar la no estacionariedad en varianza, entonces la serie IRAS requiere una transformación logarítmica para estabilizar la varianza.

4.1.2. Estacionariedad en varianza de IRAS

Figura 4.3: Gráfico de rango-media para IRAS sin transformación



En el gráfico de rango-media se han tomado 7 grupos de 12 observaciones y para cada grupo se ha calculado el rango y la media, en el eje de las abscisas se muestran las medias y en el eje de las ordenadas aparecen los rangos. Se observa que los puntos se ajustan una recta de pendiente positiva (aunque el ajuste no es muy significativo), por lo tanto se concluye que se necesita de una transformación logarítmica para estabilizar la varianza de la serie IRAS. Para verificar que la varianza es más estable se muestra el gráfico de la serie de tiempo IRAS con transformación logarítmica.

Figura 4.4: Gráfico de IRAS con logaritmos

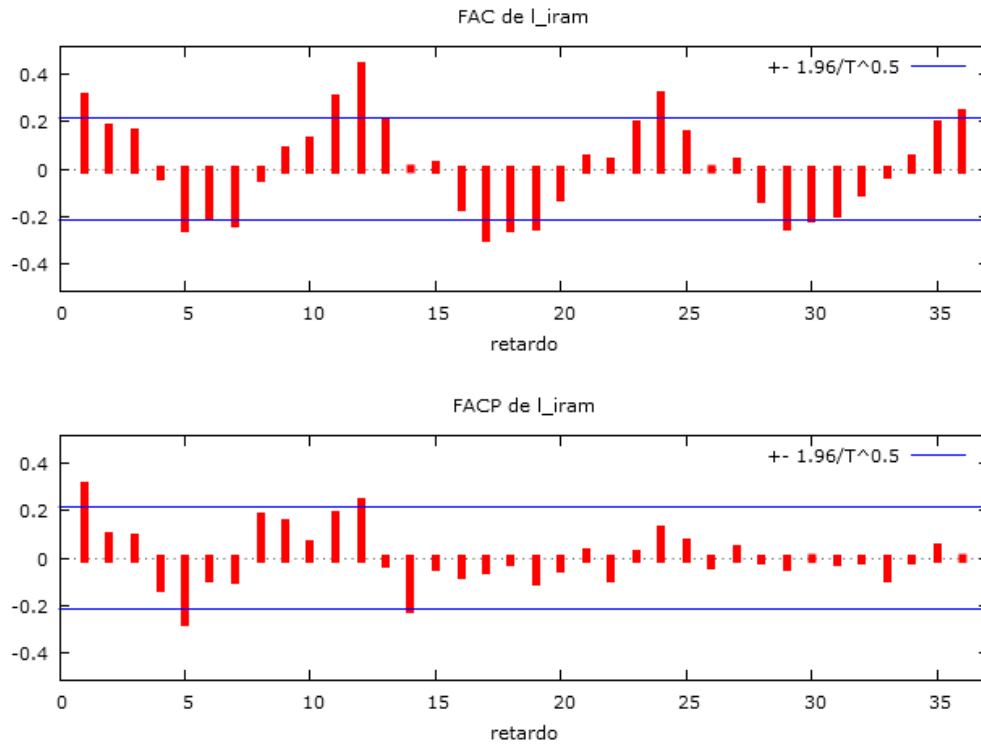


Se observa que la serie es ya estacionaria en varianza porque la variabilidad es más o menos constante para toda la serie.

4.1.3. Estacionalidad IRAS

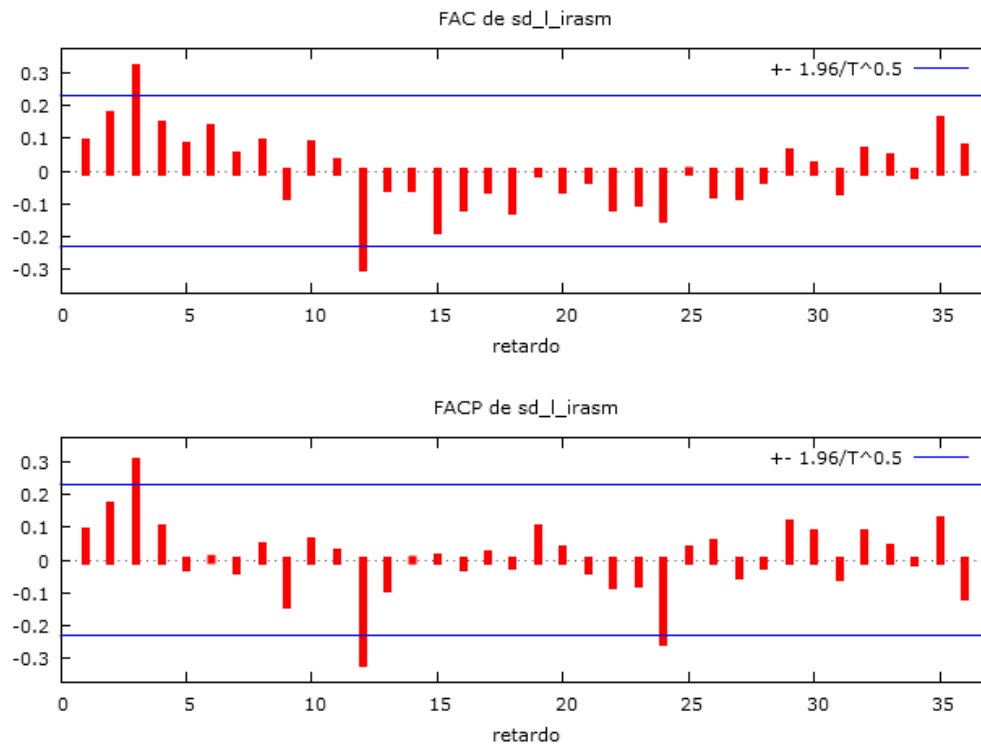
Como la serie ya es estacionaria en media y en varianza para la parte regular (primeros retardos consecutivos), ahora se procede a analizar la parte estacional. Para analizar la estacionariedad en la parte estacional hay que observar el gráfico de la serie transformada en logaritmos y su correlograma correspondiente.

Figura 4.5: Correlograma de la serie IRAS con transformación logarítmica



Se observa en el gráfico 4.4 que la serie presenta patrones estacionales, es decir, que hay un comportamiento similar en los meses de distintos años. Esto puede verificarse la figura 4.5 que muestra en los retardos 12, 24 y 36 correlaciones significativas (se salen de las bandas de significación), por lo tanto la serie de tiempo IRAS requiere de una diferenciación estacional para estabilizar la parte estacional. La figura 4.6 muestra el correlograma de dicha serie.

Figura 4.6: Correlograma de la serie IRAS con logaritmos y una diferencia estacional



El correlograma muestra solo un retardo significativo (se sale de las bandas rojas) para la FAC y en la parte regular (primeros retardos consecutivos) solo es significativa la autocorrelación para el retardo 3 y los primeros cinco retardos tienen un crecimiento amortiguado hacia cero (crecen y decrecen hacia cero de forma rápida).

En la FACP ocurre algo similar, en la parte estacional son significativos los primeros dos retardos (es decir, los retardos 12 y 24 se salen de la banda roja) y en la parte regular también hay un crecimiento amortiguado hacia cero. Por lo que se identifica un posible modelo ARIMA con media móvil de orden 1 en la parte estacional ($SMA(1)_{12}$) porque en la FAC solo se sale el primer retardo estacional y en la FACP se muestra un decrecimiento rápido hacia cero después de los primeros dos retardos estacionales y en la parte re-

gular un posible $ARMA(1, 0, 1)$ debido a que presenta una decrecimiento amortiguado hacia cero de forma similar en la FAC y en la FACP, entonces el modelo identificado es: $ARIMA(1, 0, 1)(0, 1, 1)_{12}$

4.2. Estimación para el modelo IRAS

La estimación de los parámetros del modelo se muestra en la siguiente tabla

Modelo : ARIMA, usando las observaciones Enero 2008 a

Diciembre 2013 ($T = 72$)

	Coefficiente	Desv. T.	z	Valor p
ϕ_1	0.947	0.0701	13.496	0.000
θ_1	-0.800	0.097	-8.25	0.000
Θ_1	-0.668	0.152	-4.38	0.00

Se observa que todos los parámetros o coeficientes del modelo son significativos (distintos de cero), $\phi_1 = 0,947$ es el coeficiente para un $AR(1)$ regular, $\theta_1 = -0,80$ es el coeficiente de un $MA(1)$ regular y $\Theta_1 = -0,668$ es el coeficiente de un $MA(1)_{12}$ estacional. Cada uno de los coeficientes tienen un p valor casi cero, por lo tanto los parámetros son adecuados para el modelo ARIMA. La ecuación del modelo para la series IRAS es:

$$(1 - B^{12})(1 - 0,947B)ln(IRAS) = (1 + 0,80B)(1 + 0,668B^{12})\varepsilon_t$$

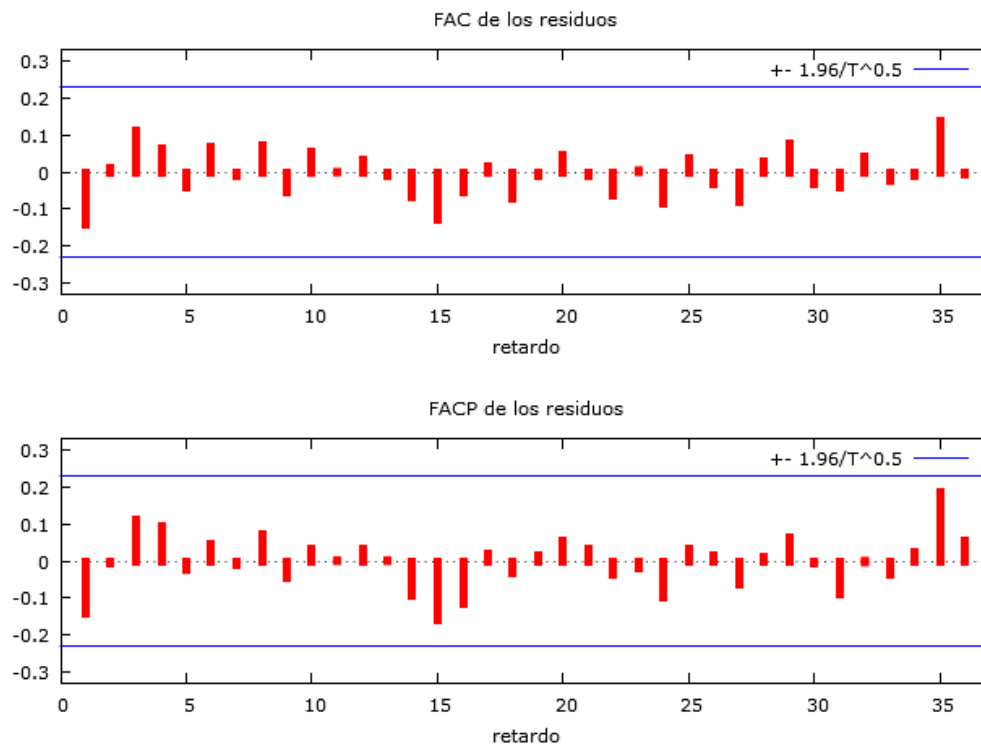
4.3. Diagnósis o Chequeo de IRAS

En esta etapa se comprueba la normalidad de los residuos del modelo y la incorrelación de los residuos, si una de estas condiciones no se cumple es porque el modelo no es adecuado.

4.3.1. Incorrelación de residuos de IRAS

En el gráfico 4.7 se muestra el correlograma de los residuos.

Figura 4.7: Correlograma de residuos



El gráfico 4.7 muestra que los residuos del modelos ARIMA identificado son incorrelados, es decir, todos o la mayoría no se salen de las bandas rojas que representan el intervalo de confianza a un nivel de confianza el 95 por ciento. Se acepta incorrelación de los residuos si como máximo el cinco por ciento de los retardos salen de las bandas azules(cinco por ciento de 36 retardos), el gráfico anterior muestra que en 36 retardos ninguno se sale del intervalo de confianza.

Para verificar lo anterior se muestra la prueba de hipótesis sobre la incorrelación de los residuos, esta prueba es el contraste LM de Breusch-

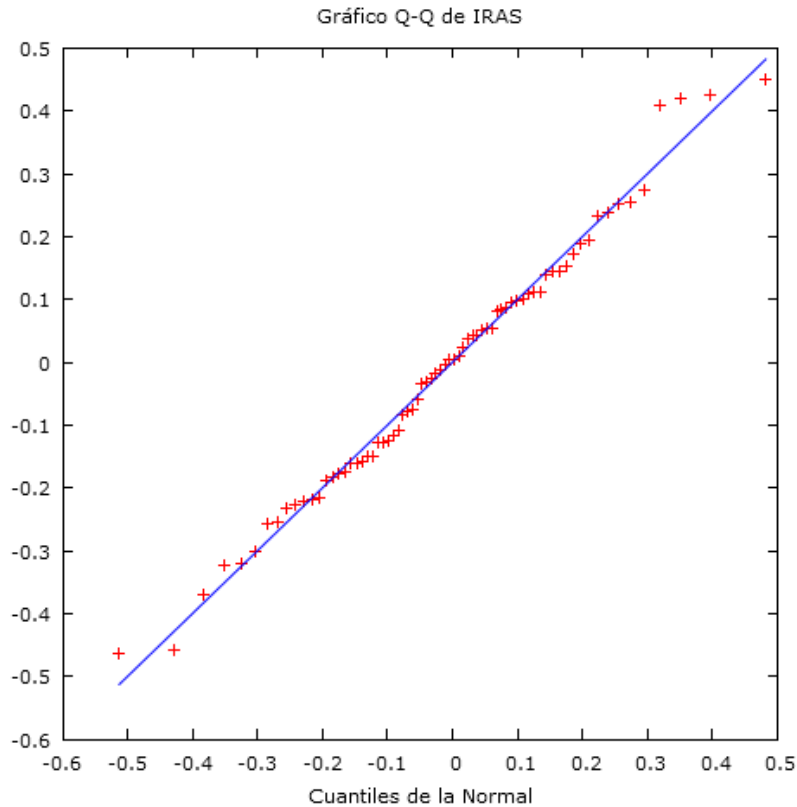
Godfrey¹⁷ con Hipótesis nula: No hay autocorrelación, que muestra un valor de $\chi^2 = 4,33824$ y un p valor igual a 0.8878. Como el p valor es mayor que el nivel de significancia de cinco por ciento entonces se acepta la hipótesis nula de que los residuos del modelo son incorrelados.

4.3.2. Normalidad de los residuos para IRAS

Para verificar la normalidad de los residuos se utiliza el estadístico Chi cuadrado con hipótesis nula: Existe normalidad, y el gráfico QQ de los residuos el cual comprueba la existencia de normalidad si los puntos del gráfico de dispersión se ajustan muy bien a la recta con pendiente positiva $y = x$. El valor del estadístico es $\chi^2 = 0,290$ con un p valor igual a **0,86488**, Como el p valor es mayor que el nivel de significancia de cinco por ciento entonces se acepta la hipótesis nula de que los residuos del modelo son normales.

¹⁷Esta prueba la hace el programa GRETL por defecto

Figura 4.8: Gráfico QQ de residuos de la serie IRAS.

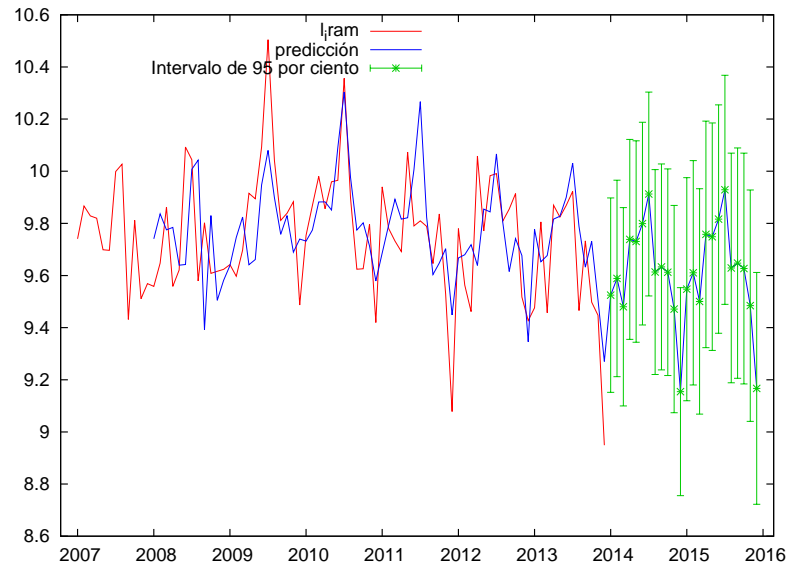


El gráfico 4.8 muestra un buen ajuste de los puntos hacia la recta $y=x$ por lo tanto hay normalidad en los residuos y el modelo pasa la etapa de Chequeo(verificación) y ya se pueden hacer predicciones con el modelo $ARIMA(1, 0, 1)(0, 1, 1)_{12}$ para la serie temporal IRAS. Las predicciones se muestran en en Anexo 2.

4.3.3. Predicciones IRAS

En el siguiente gráfico se muestran los datos graficados con las predicciones correspondientes a la serie IRAS, y se observa que los datos observados se ajustan bien a los pronósticos.

Figura 4.9: Predicciones IRAS.



En la siguiente tabla se muestran las predicciones para los años 2013-2014

Tabla 4.1: **Datos y Predicciones de casos de IRAS para los años 2013-2014**

Año-Mes	IRAS	Predicciones
2013-01	13042	17619.99
2013-02	18138	15564.86
2013-03	12801	15941.11
2013-04	19337	18332.10
2013-05	18476	18534.07
2013-06	19321	19926.03
2013-07	20385	22723.67
2013-08	12908	17819.88
2013-09	16862	15246.62

2013-10	13328	16828.12
2013-11	12674	13448.56
2013-12	7702	10614.81
2014-01		13693.83
2014-02		14600.35
2014-03		13097.47
2014-04		16957.22
2014-05		16822.63
2014-06		18016.96
2014-07		20179.57
2014-08		14963.02
2014-09		15259.64
2014-10		14953.46
2014-11		12980.36
2014-12		9456.15

La tabla 4.1 muestra que el mayor número de casos de IRAS para el año 2013 se dan en los meses: Junio y Julio, esto significa que las Infecciones Respiratorias se dan más en la época de invierno. Los pronósticos para los meses Junio y Julio del año 2014 muestran un comportamiento similar a lo que se dió en al año 2013. El número de casos de IRAS en el mes de Junio es aproximadamente 18,016 y en el mes de Julio aproximadamente 20179. Se observa que el número de casos disminuye en el año 2014 según los pronósticos. El menor número de casos se dá en Diciembre tanto para el año 2013 como para el año 2014.

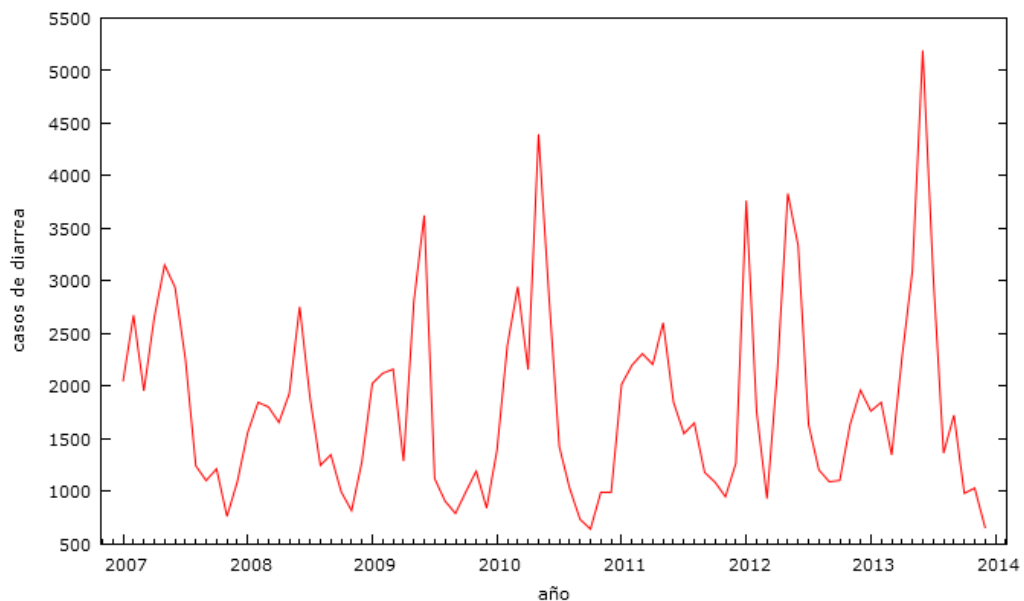
4.4. Identificación del modelo para Diarrea

La etapa de identificación se inicia con el gráfico de la serie temporal y su correspondiente correlograma, en el gráfico se trata de identificar la tendencia, estacionalidad y estacionariedad en media y en varianza.

4.4.1. Estacionariedad en media de Diarrea

En el siguiente gráfico se observa que la media es más o menos constante, es decir los datos de la serie de tiempo están alrededor de la media aritmética de los datos de la serie de tiempo Diarrea. Además la serie presenta un patrón estacional.

Figura 4.10: Gráfico de la serie Diarrea sin transformaciones desde Enero 2007 hasta Diciembre 2013.

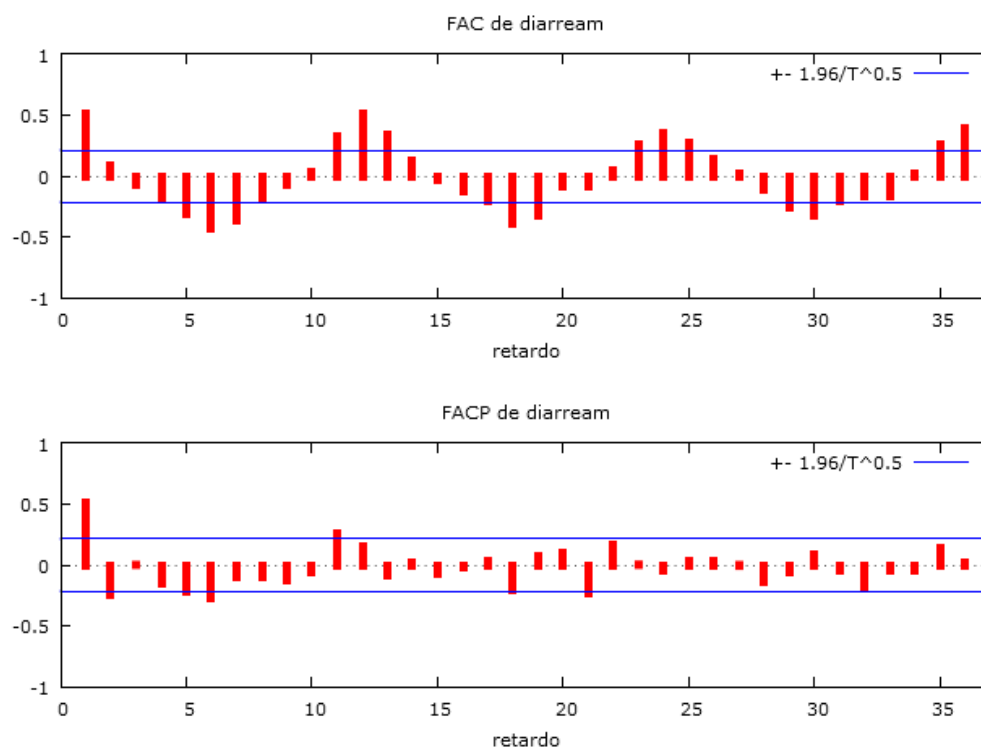


Como la serie Diarrea ya es estacionaria en media, no es necesaria diferenciarla regularmente; ésto solo se haría en caso de no ser estacionaria en media.

Otra forma de ver si existe estacionariedad en media es por medio del correlograma, la figura 4.11 confirma la estacionariedad en media ya que en

los primeros retardos las autocorrelaciones muestran un rápido decrecimiento hacia cero, aunque en la parte estacional (retardos 12, 24 y 36) la serie presenta en su FAC autocorrelaciones significativas (se salen de las bandas azules)

Figura 4.11: Correlograma de la serie Diarrea sin transformaciones



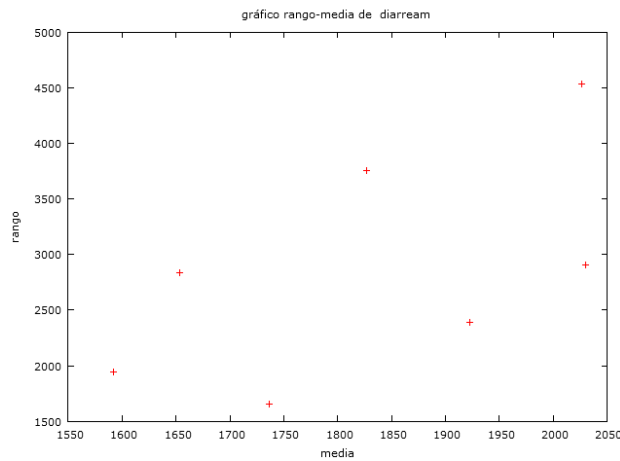
Se observa que los primeros retardos del correlograma tienen un decrecimiento hacia cero, es decir, la serie es estacionaria en media, aunque en la parte estacional (retardos 12, 24 y 36) muestra mucha significancia en las autocorrelaciones.

4.4.2. Estacionariedad en varianza de Diarrea

El gráfico de la serie Diarrea muestra que la variabilidad de la serie no es constante porque crece con el tiempo, es decir, no hay una única varianza para

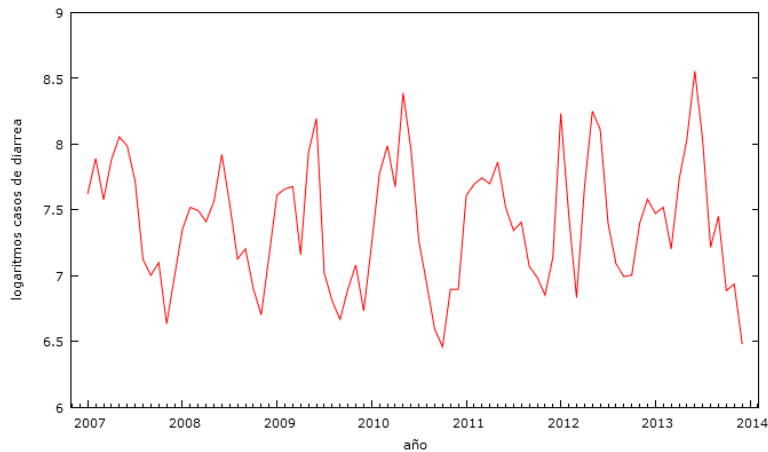
los datos de la serie temporal. Esto puede verificarse en el gráfico de rango-media que muestra 7 subgrupos (uno por año) para los cuales se ha calculado la media y el rango, en el eje de las abcisas se muestra los rangos de los 7 años y en el eje de las ordenadas están las medias por año. Como los puntos se ajustan a una recta con pendiente positiva entonces no hay estacionariedad en varianza, por lo tanto la serie Diarrea necesita una transformación logarítmica para estabilizar la varianza de la serie.

Figura 4.12: Gráfico de rango-media de la serie Diarrea



En la figura 4.13 se muestra el gráfico de la serie de tiempo Diarrea con transformación logarítmica, y se observa que dicha transformación estabiliza la varianza.

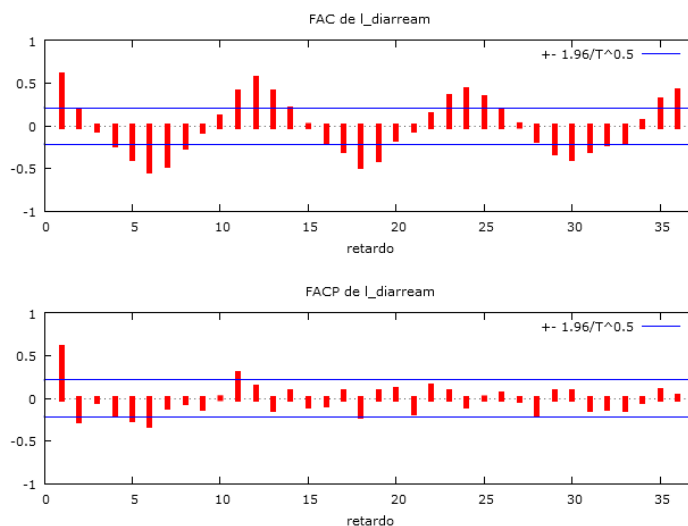
Figura 4.13: Gráfico de la serie Diarrea con transformación logarítmica



4.4.3. Estacionalidad de la serie Diarrea

En la figura 4.13 se observa que la serie de tiempo Diarrea tiene estacionalidad (patrones relacionados en periodos menores de un año), esto puede verificarse en el correlograma de la serie con transformación logarítmica que se muestra en la figura 4.14.

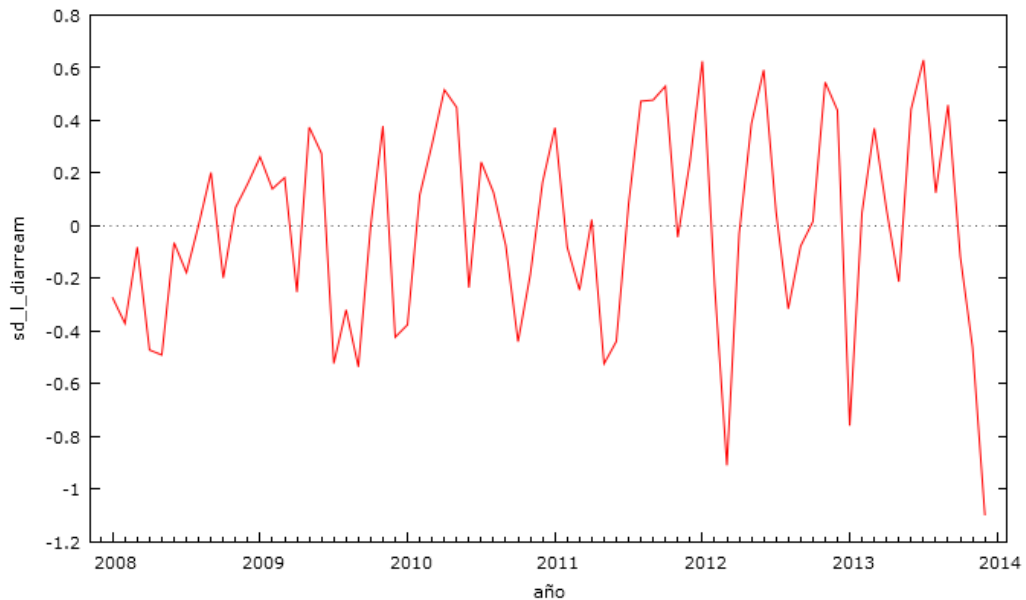
Figura 4.14: Correlograma de la serie Diarrea con logaritmos.



El correlograma muestra que no hay estacionariedad en la parte estacional porque en la FAC las autocorrelaciones para los retardos 12, 24 y 36 son

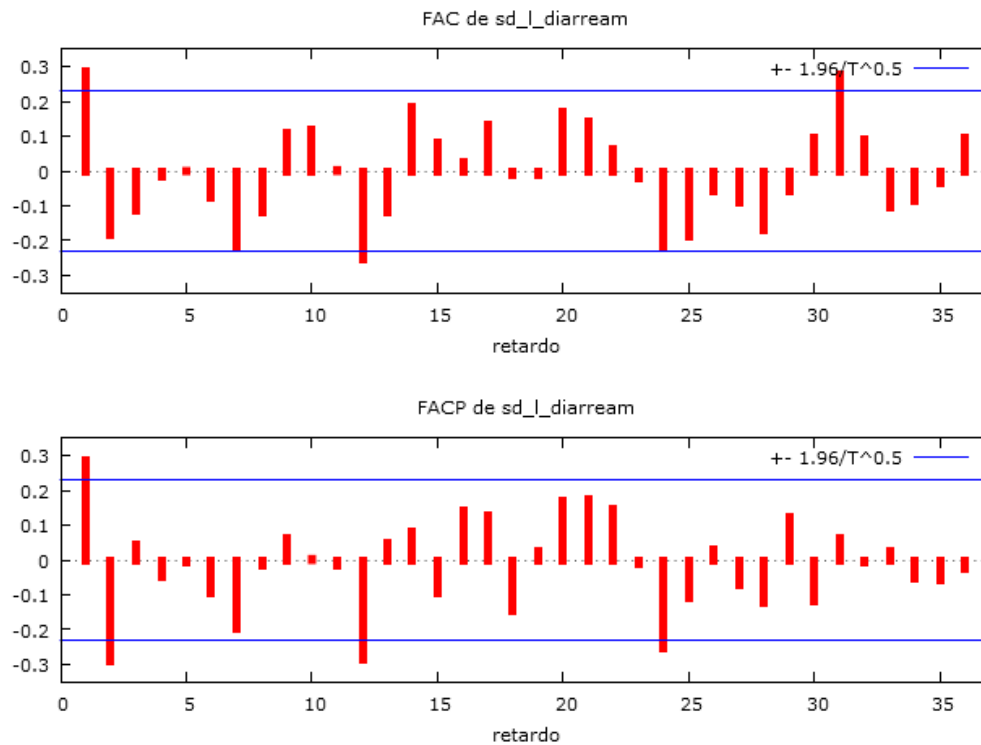
bastante significativas (se salen del intervalo de confianza) por lo tanto la serie Diarrea con transformación logarítmica debe ser diferenciada estacionalmente para eliminar el patrón estacional. A continuación se muestra el gráfico de la serie Diarrea con transformación logarítmica y una diferencia estacional.

Figura 4.15: Gráfico de la serie Diarrea con logaritmos y una diferencia estacional.



Se observa que ya no hay un patrón estacional en la serie, por lo tanto ya se puede identificar un modelo ARIMA para la serie Diarrea. Para Identificar un posible modelo la herramienta más adecuada es el correlograma de la serie, cuando ésta ya es estacionaria en media y varianza y sin estacionalidad.

Figura 4.16: correlograma de la serie Diarrea con logaritmos y una diferencia estacional



La FAC muestra un decrecimiento amortiguado hacia cero y tiene el primer retardo significativo en la parte regular, en la parte estacional muestra un decrecimiento hacia cero y es significativo el primer retardo estacional (retardo 12), la FACP en la parte regular muestra los primeros dos retardos significativos y en la parte estacional un decrecimiento hacia cero, aunque los primeros dos retardos estacionales (retardo 12 y 24) son significativos. Entonces los posibles modelos son: **ARIMA(2,0,1)(0,1,1)**, **ARIMA(1,0,1)(0,1,1)** y **ARIMA(1,0,0)(0,1,1)** en caso de no ser adecuados hay que buscar uno o varios modelos que pasen todas las etapas de la metodología Box-Jenkins, y si hay varios modelos que son adecuados se elige el que tenga menos parámetros, es decir, se utiliza el principio de **Parsimonia** que consiste en elegir el modelo más sencillo.

4.5. Estimación de la serie Diarrea.

En esta etapa se evalúa si los modelos identificados son buenos o si hay que desecharlos y seleccionar otros, a continuación se muestran los coeficientes o parámetros para el modelo $ARIMA(2, 0, 1)(0, 1, 1)$

Modelo ARIMA, usando las observaciones 2008:01-2013:12 ($T = 72$)

	Coefficiente	Desv. Típica	z	Valor p
ϕ_1	-0.388	0.294	-1.321	0.186
ϕ_2	0.0667	0.169	0.394	0.693
θ_1	0.739	0.267	2.763	0.005
Θ_1	-1.000	0.533	-1.874	0.061
Media de la vble. dep.	-0.0009	D.T. de la Var. Dep.	0.378	
media innovaciones	-0.0031	D.T. innovaciones	0.267	
Log-verosimilitud	-18.838	Criterio de Akaike	47.676	
Criterio de Schwarz	59.059	Hannan-Quinn	52.208	

Se observa que los coeficientes autorregresivos(AR), ϕ_1 y ϕ_2 no ayudan a obtener un buen modelo porque tienen un valor de z menor que 2 y un p valor mayor que 0.05, así también Θ_1 tiene un coeficiente de -1.0 (los coeficientes deben estar entre -1 y 1) lo cual indica que el MA estacional no es bueno para explicar el modelo. Como el el modelo $ARIMA(2, 0, 1)(0, 1, 1)$ no es adecuado, probamos el segundo modelo, éste es el $ARIMA(1, 0, 1)(0, 1, 1)$, a continuación se muestra sus parámetros con sus respectivos z , Desviaciones Típicas y su p valor.

Modelo ARIMA, usando las observaciones 2008:01-2013:12

	Coefficiente	Desv. Típica	z	Valor p
ϕ_1	-0.310888	0.315289	-0.9860	0.3241
θ_1	0.650988	0.267689	2.4319	0.0150
Θ_1	-1.00000	0.638622	-1.5659	0.1174
Media de la vble. dep.	-0.0009	D.T. de la vble. dep.		0.378
media innovaciones	-0.0026	D.T. innovaciones		0.267
Log-verosimilitud	-18.912	Criterio de Akaike		45.825
Criterio de Schwarz	54.932	Hannan-Quinn		49.451

Se observa que ϕ_1 y Θ_1 tienen un z menor que 2, por lo tanto este modelo no es adecuado porque no pasa la etapa de estimación, entonces hay que probar con el modelo $ARIMA(1, 0, 0)(0, 1, 1)_{12}$

Modelo ARIMA, usando las observaciones 2008:01-2013:12

	Coefficiente	Desv. Típica	z	Valor p
θ_1	0.363590	0.127861	2.8436	0.0045
Θ_1	-1.00000	0.676619	-1.4779	0.1394
Media de la vble. dep.	-0.0009	D.T. de la vble. dep.		0.378
media innovaciones	-0.003	D.T. innovaciones		0.268
Log-verosimilitud	-19.241	Criterio de Akaike		44.483
Criterio de Schwarz	51.313	Hannan-Quinn		47.202

Se observa que θ_1 es significativo porque tiene un valor de z mayor que 2 y Θ_1 no es significativo para el modelo porque tiene un z menor que 2, por lo tanto hay que quitar a Θ_1 del modelo, a continuación se muestran las estimaciones del nuevo modelo que sería $ARIMA(0, 0, 1)(0, 1, 0)$ La ecuación del modelo para la serie Diarrea es:

$$(1 - B^{12})\ln(Diarrea) = (1 - 0,529B)\varepsilon_t$$

Modelo ARIMA, usando las observaciones 2008:01–2013:12

	Coefficiente	Desv. Típica	z	Valor p
θ_1	0.529539	0.114881	4.6095	0.0000
Media de la vble. dep.	-0.00097	D.T. de la vble. dep.	0.378	
media innovaciones	-0.0056	D.T. innovaciones	0.342	
Log-verosimilitud	-25.247	Criterio de Akaike	54.495	
Criterio de Schwarz	59.048	Hannan–Quinn	56.308	

Se observa que este modelo si es adecuado porque el coeficiente θ_1 tiene valor z mayor que 2, entonces este modelo pasa la etapa de estimación.

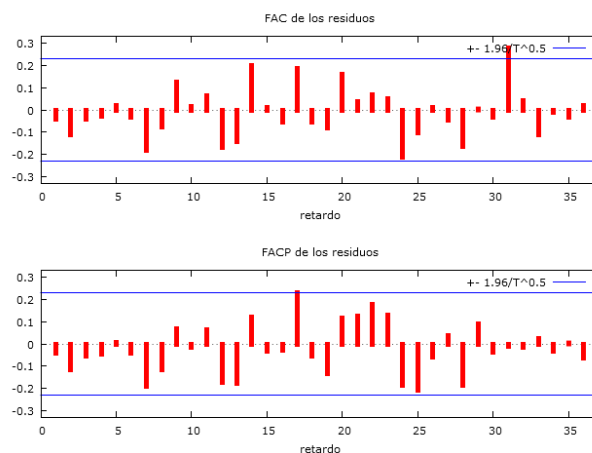
4.6. Diagnósis o Chequeo de Diarrea

En esta etapa se estudia la incorrelación de los reídos y su normalidad, a continuación se muestra el correlograma de los resíduos para verificar que tenga media cero y varianza constante(ruido blanco).

4.6.1. Incorrelación de resíduos de Diarrea

En el grafico 4.17 se muestra el correlograma de los resíduos.

Figura 4.17: Correlograma de resíduos de la serie Diarrea.



Se observa que la mayoría de resíduos están en el intervalo de confianza (bandas azules) y los primeros resíduos son casi cero, en la FAC solo se

sale del intervalo de confianza el residuo del retardo 32, por lo tanto solo se salen 1 de 36 (menos del cinco por ciento)resíduos, por lo tanto los resíduos de la serie son incorrelados, es decir, no están relacionados. Para verificar lo anterior se muestra la prueba de hipotesis sobre la incorrelación de los resíduos, esta prueba es el contraste Ljung-Box Q' ¹⁸ con Hipótesis nula: No hay autocorrelación, que muestra un valor de $Q' = 8,87244$ y un p valor igual a 0.6337. Como el p valor es mayor que el nivel de significancia de cinco por ciento entonces se acepta la hipótesis nula de que los resíduos del modelo son incorrelados.

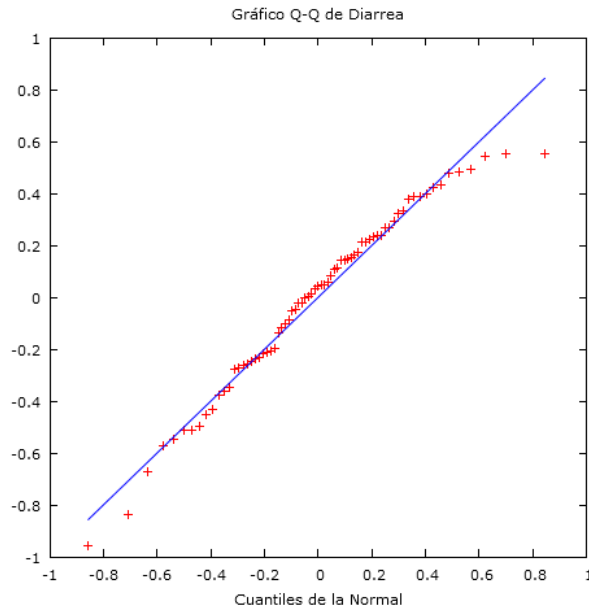
4.6.2. Normalidad de los residuos para Diarrea

Para ver si hay normalidad de los residuos se utiliza el estadístico Chi cuadrado con hipótesis nula: Existe normalidad, y el gráfico QQ de los resíduos el cual comprueba la existencia de normalidad si los puntos del gráfico de dispersión se ajustan muy bien a la recta con pendiente positiva $y = x$. El valor del estadístico es $\chi^2 = 4,005$ con un p valor igual a 0.13503, como el p valor es mayor que el nivel de significancia de cinco por ciento entonces se acepta la hipótesis nula de que los resíduos del modelo son normales.

En la siguiente figura se muestra el gráfico QQ para comprobar si hay normalidad en los resíduos.

¹⁸Esta prueba la hace el programa GRETl por defecto

Figura 4.18: Gráfico QQ de residuos de la serie Diarrea.



En este gráfico todos los puntos se ajustan bastante bien a la recta $y = x$ por lo tanto los residuos son normales. Como el modelo pasa todas las etapas de la metodología Box-Jenkins, entonces ya se puede predecir el número de casos de Diarrea para los meses de los años 2014 y 2015, las predicciones se muestran en el ANEXO 3.

4.6.3. Predicciones de la serie Diarrea

Figura 4.19: Gráfico de la serie de tiempo Diarrea y sus Predicciones

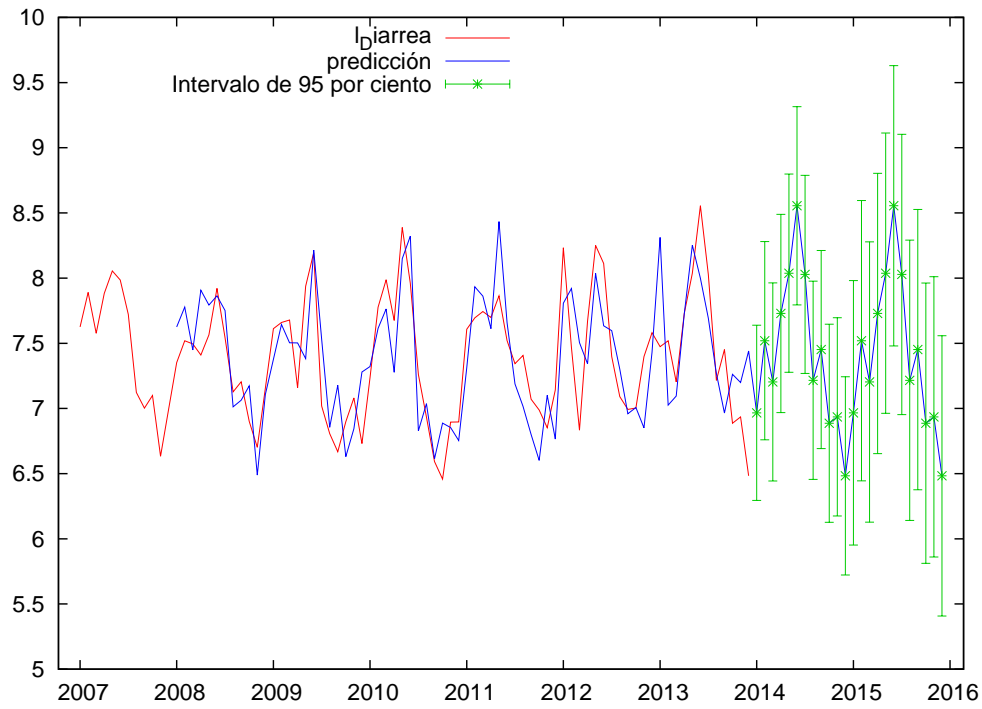


Tabla 4.3: Datos y Predicciones de casos de Diarrea para los años 2013-2014

2013-01	1761	4073.511
2013-02	1844	1125.036
2013-03	1344	1205.546
2013-04	2273	2265.756
2013-05	3095	3837.481
2013-06	5193	2976.08

2013-07	3067	2192.879
2013-08	1361	1434.486
2013-09	1723	1059.093
2013-10	979	1424.642
2013-11	1028	1337.969
2013-12	654	1704.703
2014-01		1060.312
2014-02		1844
2014-03		1344
2014-04		2273
2014-05		3095
2014-06		5193
2014-07		3067
2014-08		1361
2014-09		1723
2014-10		979
2014-11		1028
2014-12		654

Datos y Pronósticos de Diarrea año 2013												
Mes	En.	Feb.	Marz.	Ab.	May.	Jun.	Jul.	Ag.	Sep.	Oct.	Nov.	Dic.
Datos	1761	1844	1344	2273	3095	5193	3067	1361	1723	979	1028	654
Pronósticos	4073.511	1125.036	1205.546	2265.756	3837.481	2976.08	2192.879	1434.486	1059.093	1424.642	1337.969	1704.703
Datos y Pronósticos de Diarrea año 2014												
Mes	En.	Feb.	Marz.	Ab.	May.	Jun.	Jul.	Ag.	Sep.	Oct.	Nov.	Dic.
Datos												
Pronósticos	1060.312	1844	1344	2273	3095	5193	3067	1361	1723	979	1028	654

La tabla anterior muestra que los meses en los que los números de casos de Diarrea aumentan son desde Mayo hasta Julio tanto para el año 2013 como para el año 2014. Esto se debe a que en la época de invierno hay más humedad, más frutas, y es la época donde hay abundancia de moscas. Se puede observar también que el error (dato observado menos la predicción) es bastante grande en la mayoría de meses.

Cada pronóstico se interpreta como el número de casos de Diarrea que se esperan según el modelo ARIMA identificado, los pronósticos pueden ser más grandes o pequeños que el dato observado porque no son exactos.

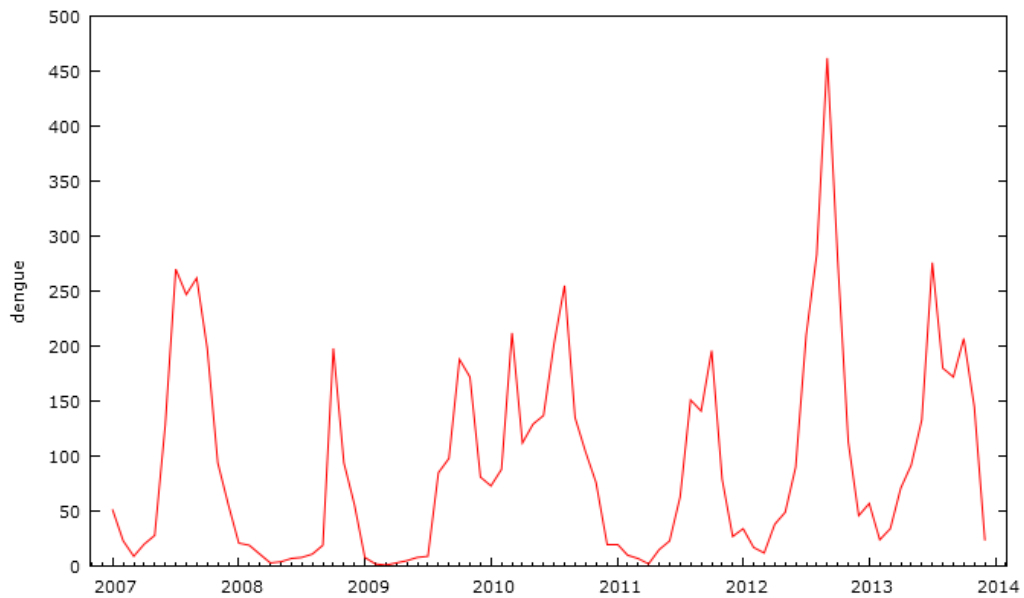
4.7. Identificación del modelo para Dengue

Para la etapa de identificación se gráfica la serie de tiempo para observar estacionariedad en media y en varianza, así como también observar patrones de tendencia y estacionalidad.

4.7.1. Estacionariedad en media para Dengue

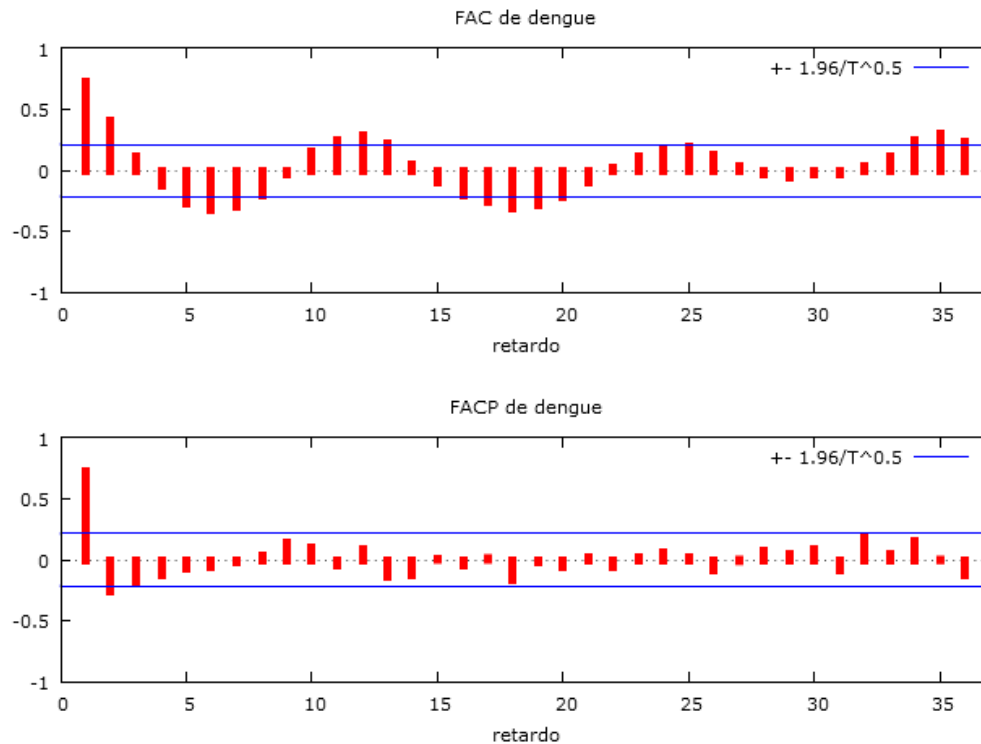
Para analizar la estacionariedad en media hay que observar el gráfico de la serie de tiempo y ver si los datos están alrededor de la media de la serie, es decir, se trata de que no hayan varias medias o varios tramos con medias distintas. En la figura 4.20 se muestra el gráfico de la serie Dengue y observa que es más o menos estacionario en media porque los datos están en torno a una única media, aunque se observa un marcado componente estacional.

Figura 4.20: Gráfico de la serie de tiempo Dengue sin transformaciones desde Enero 2007 a Diciembre 2013.



Para confirmar la estacionariedad en media a continuación se muestra el correlograma de la serie Dengue sin transformaciones.

Figura 4.21: Correlograma de la serie Dengue con transformaciones.

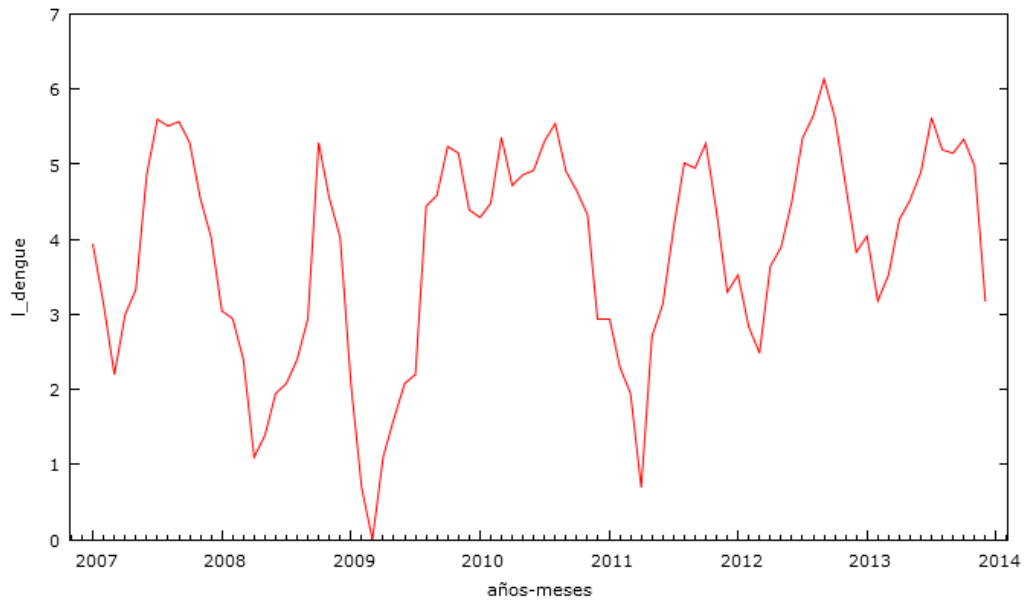


La figura 4.21 muestra que la serie de tiempo Dengue es estacionaria en media porque en la FAC y en la FACP solo los primeros dos retardos son significativos (se salen de las bandas azules) y hay un decrecimiento hacia rápido hacia cero en los primeros retardos consecutivos.

4.7.2. Estacionariedad en varianza de Dengue

La estacionariedad en varianza consiste en identificar una única varianza para los datos de la serie, es decir, que los datos tengan variabilidad estable. En la figura 4.20 se observa que la variabilidad no es estacionaria o estable porque entre 2012 y 2013 hay un pico muy alto, eso significa que la variabilidad es diferente en ese tiempo, por lo tanto la serie debe ser transformada en logaritmos para lograr la estacionariedad en varianza. En la figura 4.22 se presenta el gráfico de la serie Dengue con transformación logarítmica.

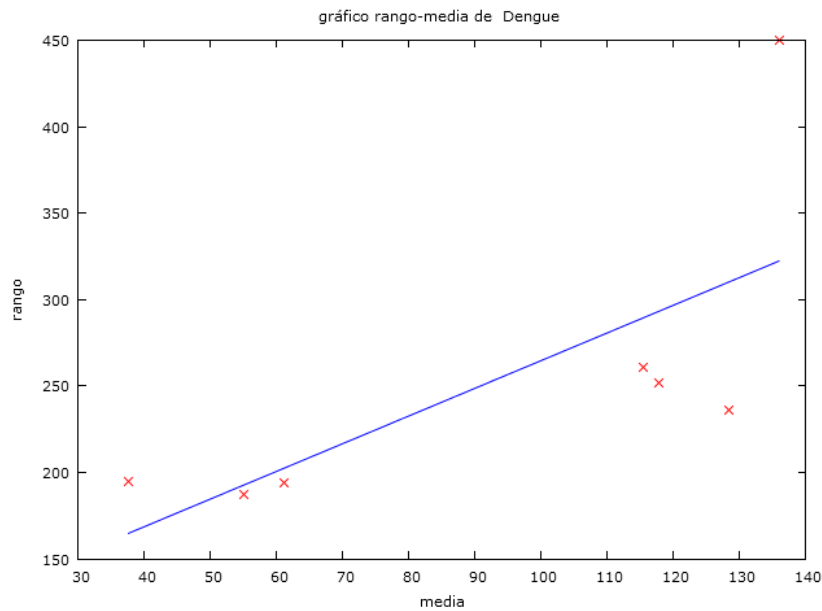
Figura 4.22: Gráfico de la serie Dengue con logaritmos , Enero 2007-Diciembre 2013.



Se observa que la serie con logaritmos es más o menos estable en varianza, porque ya no tiene picos muy pronunciados como en la figura 4.20. Otra forma de identificar la estacionariedad o no estacionariedad en varianza es a partir del gráfico de rango-media el cual en el eje Y muestra el rango¹⁹ de la serie por cada año y en el eje X se muestra las correspondiente media por cada año, si los puntos se ajustan a una recta con pendiente positiva (aunque el ajuste no sea muy bueno) entonces la serie de tiempo no es estacionaria en varianza, en la figura 4.23 se muestra el gráfico de rango-media.

¹⁹Rango: Dato mayor menos dato menor

Figura 4.23: Gráfico de rango-media de la serie Dengue

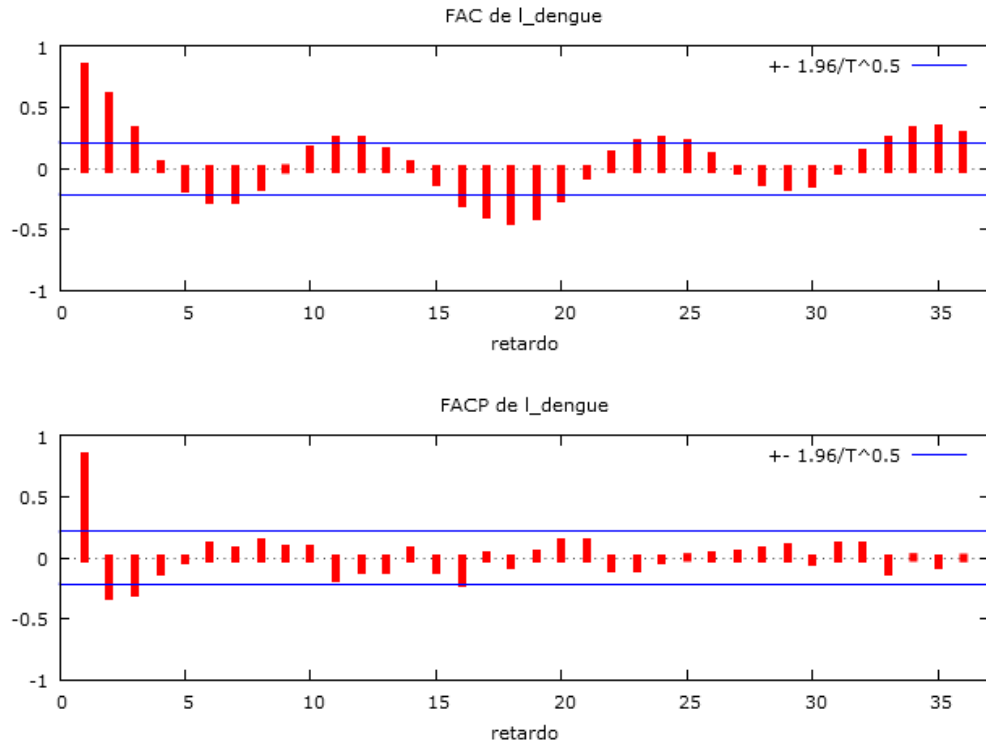


Se observa que los puntos se ajustan a una recta con pendiente positiva, entonces hay que aplicar una transformación logarítmica a los datos de la serie Dengue. El gráfico de la serie con transformación logarítmica se muestra en la figura 4.22.

4.7.3. Estacionalidad de la serie Dengue

En la figura 4.20 se observa un marcado patrón estacional en la serie Dengue, lo cual sugiere una diferenciación estacional. Otra forma de ver la estacionalidad es con el Correlograma de la serie de tiempo, en la figura 4.24 se muestra el correlograma de la serie de tiempo Dengue con transformación logarítmica.

Figura 4.24: Correlograma de la serie Dengue con logaritmos



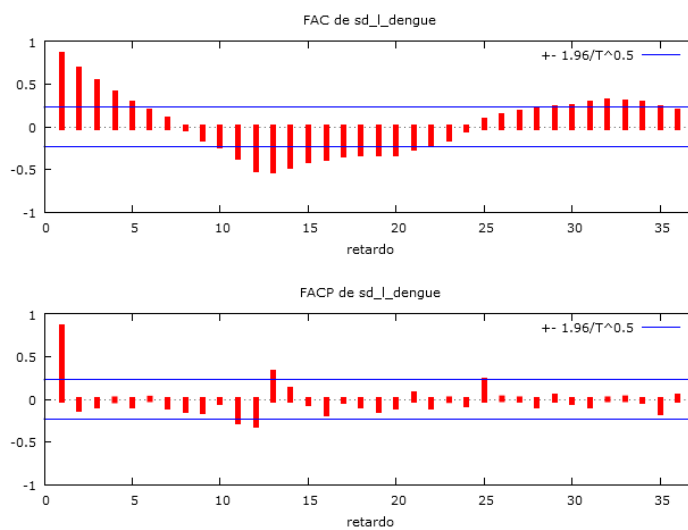
Se observa que las correlaciones en los retardos estacionales (retardos 12, 24 y 36) son significativas en la FAC, por lo tanto se sugiere una diferencia estacional para eliminar el patrón estacional y lograr estacionariedad estacional. En la siguiente figura se muestra el gráfico de la serie de tiempo Dengue con logaritmos y una diferencia estacional.

Figura 4.25: Gráfico de la serie Dengue con logaritmos y una diferencia estacional.



Se observa que ya no hay un patrón estacional, por lo tanto se procede a identificar un modelo ARIMA para la serie Dengue. En la figura 4.26 se muestra el correlograma de la serie Dengue con logaritmos y una diferencia estacional para identificar un modelo adecuado para la serie Dengue.

Figura 4.26: Correlograma de la serie Dengue con logaritmos y una diferencia estacional.



En la FACP se observa que solo el primer retardo es significativo en la parte regular (hay un posible $AR(1)$ regular) y en la parte estacional son significativos el primer retardo (el retardo 12), en la FAC es significativo el primer retardo estacional, entonces un posible modelo es: $ARIMA(1, 0, 0)(1, 1, 0)_{12}$. A continuación se analiza la etapa de Estimación para saber si los parámetros del modelo son adecuados.

4.8. Estimación Dengue

En esta etapa se analiza si los coeficientes del modelo son importantes o no para ayudar a explicar el modelo. A continuación se muestra una tabla con los coeficientes, sus desviaciones típicas, el estadístico z y su p valor.

Modelo ARIMA, usando las observaciones 2008:01–2013:12

	Coefficiente	Desv. Tipica	z	Valor p
ϕ_1	0.837	0.06327	13.231	0.000
Φ_1	-0.5137	0.0995	-5.1602	0.0000
Media de la vble. dep.	0.0401	D.T. de la vble. dep.	1.775	
media innovaciones	-0.0041	D.T. innovaciones	0.796	
Log-verosimilitud	-88.15329	Criterio de Akaike	182.306	
Criterio de Schwarz	189.1366	Hannan–Quinn	185.025	

Se observa en la tabla que los valores de z son mayores que 2, por lo tanto los coeficientes o parámetros son adecuados para el modelo y el modelo $ARIMA(1, 0, 0)(1, 1, 0)_{12}$ pasa la etapa de estimación. Y la ecuación del modelos para la serie Dengue es:

$$(1 - B^{12})(1 - 0,837B)(1 + 0,5137B^{12}) = \varepsilon_t$$

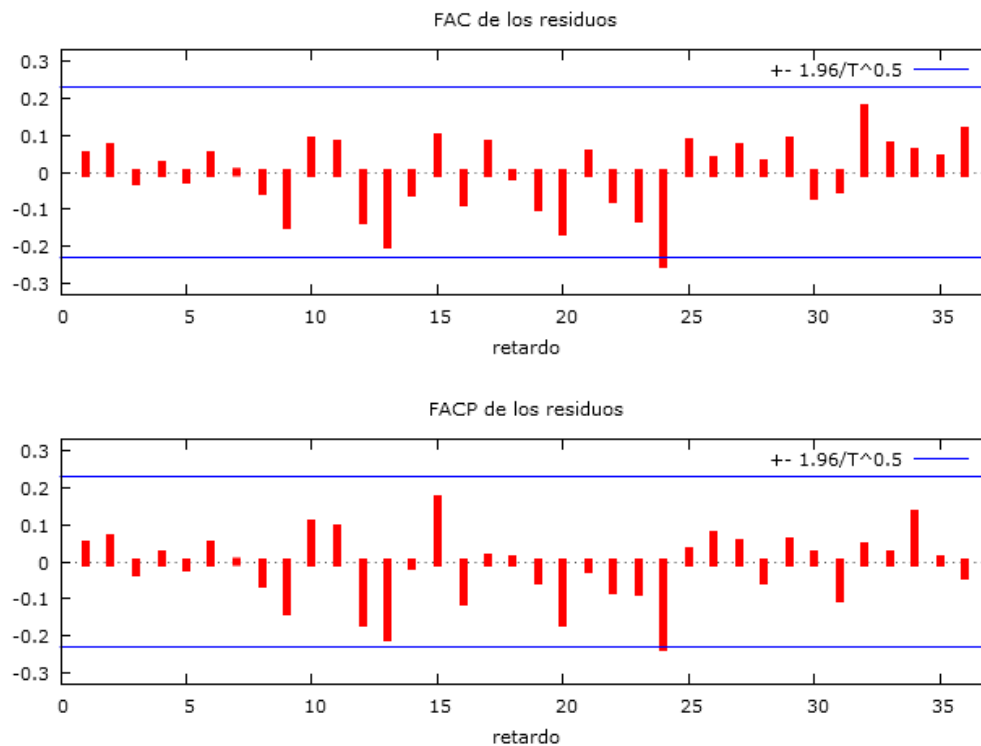
4.9. Diagnósis o Chequeo de Dengue

En esta etapa se estudia la incorrelación de los reídos y su normalidad, a continuación se muestra el correlograma de los resídos para verificar que tengan media cero y varianza constante(ruido blanco).

4.9.1. Incorrelación de los resídos de Dengue

En el grafico 4.27 se muestra el correlograma de los resídos.

Figura 4.27: Correlograma de los resídos de la serie Dengue



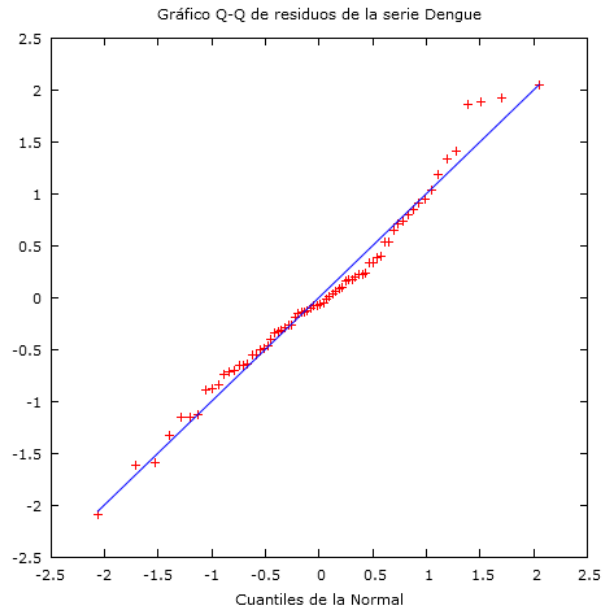
En la figura 4.27 se observa que la mayoría de resídos con casi cero, aunque del retardo 24 es significativo tanto en la FAC como en la FACP. También se observa que la mayoría de resídos están dentro de la banda lo cual confirma que los resídos son incorrelados y la mayoría están dentro de las bandas rojas; si saliera un cinco por ciento de los retardos o más, el modelo no sería adecuado porque se está trabajando con un nivel de significancia del

cinco por ciento (el 5 por ciento de 36 es aproximadamente 3, tendrían que salirse de las bandas rojas como máximo 3 retardos). Como la significación del retardo 24 en el correlograma es importante por ser un retardo estacional es posible identificar otro modelo ARIMA que puede ser un $AR(2)$ para la parte estacional, se hace mención de un $AR(2)$ porque en el modelo que se ha identificado anteriormente ya hay un $AR(1)$ en la parte estacional por lo tanto si hay otro retardo significativo estacionalmente en la FACP o en la FAC de los residuos se tiene que agregar al modelo antes identificado y éste sería un nuevo modelo ARIMA identificado como $ARIMA(1, 0, 0)(2, 1, 0)_{12}$ al cual se analizará en la próxima sección para compararlo con el modelo $ARIMA(1, 0, 0)(1, 1, 0)_{12}$ y ver cual es más parsimonioso (más sencillo y con menos parámetros), aunque puede darse el caso de que en una serie de tiempo varios modelos sean adecuados para predecir.

4.9.2. Normalidad de los residuos de Dengue

Para ver si hay normalidad de los residuos se utiliza el estadístico Chi cuadrado con hipótesis nula: Existe normalidad, y el gráfico QQ de los residuos el cual comprueba la existencia de normalidad si los puntos del gráfico de dispersión se ajustan muy bien a la recta con pendiente positiva $y = x$. El valor del estadístico es $\chi^2 = 2,04009$ con un p valor igual a 0.3605, como el p valor es mayor que el nivel de significancia de cinco por ciento entonces se acepta la hipótesis nula de que los residuos del modelo son normales. En la siguiente figura se muestra el gráfico QQ para comprobar si hay normalidad en los residuos.

Figura 4.28: Gráfico QQ de los residuos de la serie Dengue



En este gráfico todos los puntos se ajustan bastante bien a la recta $y = x$ por lo tanto los residuos son normales. Como el modelo pasa todas las etapas de la metodología Box-Jenkins, entonces ya se puede predecir el número de casos de Dengue para los meses de los años 2014 y 2015, las predicciones se muestran en el ANEXO 4.

Figura 4.29: Gráfico de la serie y pronósticos de serie Dengue

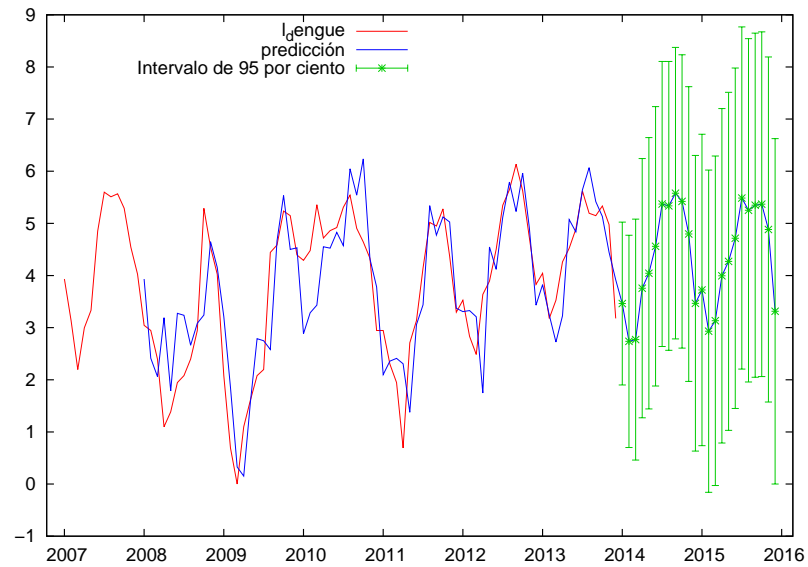


Tabla 4.4: Datos y pronósticos de la serie Dengue para los años 2013-2014

Año-Mes	Dengue	Predicciones
2013-01	57	45.815
2013-02	24	25.6231
2013-03	34	15.2555
2013-04	71	25.2428
2013-05	92	159.7411
2013-06	133	125.8988
2013-07	276	282.1297
2013-08	180	432.4285
2013-09	172	225.2399
2013-10	207	169.2413

2013-11	145	85.3639
2013-12	24	50.2808
2014-01		31.8832
2014-02		15.4364
2014-03		15.9604
2014-04		42.7921
2014-05		57.0025
2014-06		95.5744
2014-07		215.1472
2014-08		207.3628
2014-09		264.8053
2014-10		225.9851
2014-11		120.9323
2014-12		32.0605

La tabla anterior muestra que el mayor número de casos de Dengue para el año 2013 y para el 2014 se dan entre Mayo y Septiembre, debido a la época más lluviosa la cual genera más estancamientos de agua y por lo tanto hay más zancudos. Se espera que los meses entre Mayo y Junio el número de casos de Dengue para el año 2014 sea de: 57.0025, 95.5744, 215.1472, 207.3628 y 264.8053 respectivamente, se observa que en el año 2014 el número de casos de Dengue es menor con respecto al año 2013, pero esto es sólo una estimación que puede ser mayor o menor que los datos observados en la realidad.

4.10. Análisis del segundo modelo de la serie Dengue

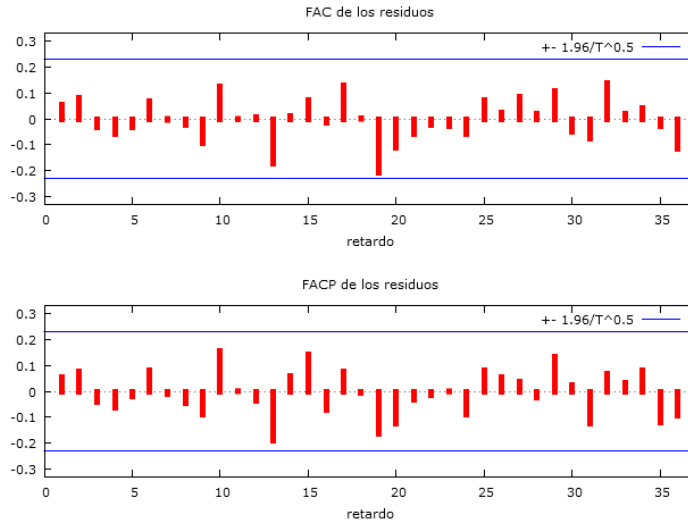
El segundo modelo identificado para la serie Dengue es el modelo **ARIMA(1,1,0)(2,1,0)** el cual se dedujo del correlograma de residuos del modelo **ARIMA(1,0,0)(1,1,0)**. A continuación analizaremos si los parámetros de este modelo **ARIMA(1,0,0)(2,1,0)** son buenos para explicarlo.

Modelo 2: ARIMA, usando las observaciones 2008:01-2013:12

	Coefficiente	Desv. Típica	z	Valor p
ϕ_1	0.805756	0.0696753	11.5644	0.0000
Φ_1	-0.752102	0.111699	-6.7333	0.0000
Φ_2	-0.462110	0.115959	-3.9851	0.0001
Media de la vble. dep.	0.0401	D.T. de la vble. dep.	1.775	
media innovaciones	-0.0073	D.T. innovaciones	0.705332	
Log-verosimilitud	-82.222	Criterio de Akaike	172.4452	
Criterio de Schwarz	181.551	Hannan-Quin	176.0706	

Se observa que los coeficiente del modelo son adecuados, es decir, tiene valores z mayores que 2. Por lo tanto este modelo pasa a la etapa de Diagnósis o Chequeo (llamada también Verificación). En la figura 4.30 se muestra el correlograma de residuos.

Figura 4.30: Correlograma de residuos para el segundo modelo de Dengue



Se puede observar que las autocorrelaciones están dentro de las bandas rojas (intervalo de confianza), y la mayoría de residuos tienen valores cercanos a cero, entonces se puede decir que los residuos son incorrelados. Para verificar si hay normalidad de los residuos se utiliza el estadístico Chi cuadrado con hipótesis nula: Existe normalidad. El valor del estadístico es $\chi^2 = 2,412$ con un p valor igual a 0.29935, como el p valor es mayor que el nivel de significancia de cinco por ciento entonces se acepta la hipótesis nula de que los residuos del modelo son normales. Por lo tanto este modelo también es bueno para predecir la serie de tiempo Dengue.

Para saber cual de los dos modelos es el mejor, se utiliza el criterio de Información de Bayes (BIC) al cual también Gretl lo calcula, el modelo que tenga el menor BIC es el mejor. El BIC para el modelo $ARIMA(1, 0, 0)(1, 1, 0)_{12}$ resulta **189,13658** y el BIC para el modelo $ARIMA(1, 0, 0)(2, 1, 0)_{12}$ resulta **181,551846**, por lo tanto se puede decir que el mejor modelo es el $ARIMA(1, 0, 0)(2, 1, 0)_{12}$ ya que tiene menor BIC, pero como la diferencia entre los BIC de ambos modelos es pequeña entonces cualquiera de los dos modelos se pueden utilizar para predecir. En este trabajo se ha utilizado

el modelo $ARIMA(1, 0, 0)(1, 1, 0)_{12}$ para predecir el número de caso de Dengue que se registrarán en los años 2014 y 2015, esto se muestra en el anexo 4.

Conclusiones

- 1 La modelación ARIMA es uno de los métodos modernos de predicción más sofisticados para series temporales con tendencia y estacionalidad. Resulta muy útil para elaborar modelos que describan adecuadamente el comportamiento del número de casos de las enfermedades que les permitan a las autoridades sanitarias estar preparadas ante cualquier eventualidad.
- 2 Es importante desarrollar modelos de predicción locales utilizando diferentes variables disponibles para ser utilizados por las comunidades en su vigilancia y su alerta epidemiológica, además de insertar en sus políticas nacionales, temas de investigación prioritarias en el control de enfermedades.
- 3 Las series temporales analizadas son estacionarias en media, pero no en varianza, lo cual significa que hay periodos de tiempo en los que el número de casos de Dengue, Diarrea e IRAS aumentan o varían demasiado con respecto a los demás meses.
- 4 Todas las series temporales analizadas tienen un patrón sistemático que se repite todos los inviernos en cada año, es decir, que el número de casos de Dengue, Diarrea e IRAS son estacionales (dependen de la estación del año)

Recomendaciones

1. La utilización de esta metodología hace posible el perfeccionamiento de la vigilancia de estas enfermedades. Aunque la modelación ARIMA es compleja y necesita de conocimientos técnicos específicos y un soporte tecnológico adecuado, en las Unidades de Análisis y Tendencias en Salud del país existen actualmente las condiciones para que esta se convierta en una de las principales herramientas para la vigilancia de problemas de salud y se recomienda su generalización.
2. Se sugiere a la población del Departamento de Santa Ana tomar las medidas necesarias que dicta el Ministerio de Salud para evitar que se den más casos de Dengue, Diarrea e Infecciones Respiratorias Agudas Superiores(IRAS).

Bibliografía

- [1] DANIEL PEÑA SÁNCHEZ DE RIVERA, *Análisis de series temporales*, Alianza Editorial, S.A, Madrid, 2005.
- [2] JONATHAN D. CRYER, KUNG-SIK CHANG , *Time Series Analysis: With applications in R*, Springer Editorial, Segunda Edición, 2008.
- [3] ROBERT H. SHUMWAY, DAVID S. STOFFER , *Time Series Analysis and Its Applications*, Springer Editorial, Tercera Edición, 2011.
- [4] J. ARNAU GRAS, *Diseño de series temporales: técnicas de análisis*, Edicions, UNIVERSITAT DE BARCELONA, España, 2001.
- [5] MARÍA PILAR GONZÁLEZ CASIMIRO, *Análisis de series temporales: Modelos ARIMA*, Departamento de Economía Aplicada III (Econometría y Estadística), Facultad de Ciencias Económicas y Empresariales Universidad del País Vasco (UPV-EHU).
- [6] *Código de salud*, Tomo nº 229, San Salvador, Miércoles 11 de Mayo de 1988, Número 86
- [7] *REGLAMENTO DE LA LEY DEL SISTEMA BÁSICO DE SALUD INTEGRAL*, DECRETO No.77.
- [8] CÉSAR PÉREZ LÓPEZ. *Econometría de las series temporales*, Editorial Pearson Prentice Hall, S.A, Madrid, 2006.
- [9] PAUL S.P COWPERTWAIT, ANDREW V. METCALFE . *Introductory Time Series with R*, Editorial Springer, USA, 2011.

- [10] JOSÉ JUAN CÁCERES HERNÁNDEZ, GLORIA MARTÍN RODRÍGUEZ, FRANCISCO JAVIER MARTÍN ÁLVAREZ . *Introducción al análisis univariante de series temporales económicas*, Editorial Delta Publicaciones, Madrid, España, 2008.
- [11] CÉSAR PÉREZ . *Técnicas Estadísticas Predictivas con IBM SPSS. Modelos*, Garceta grupo Editorial , España, 2014.
- [12] JOHN E. HANKE, ARTHUR G. REITSCH. *Pronósticos en los negocios*, Quinta Edición, Editorial Prentice Hall , México, 1996.
- [13] EDUARDO COURT MONTEVERDE, ERICK WILLIAMS RENGIFO. *Estadísticas y Econometría Financiera*. Primera Edición, Editorial CENGAGE learning, Argentina, 2011.
- [14] RAMÓN ANTONIO ROSALES ÁLVAREZ, *Fundamentos de Econometría Intermedia. Teoría y Aplicaciones*, Primera Edición, Ediciones Unian-des, Colombia, 2013.
- [15] J. BERNARDO PENA TRAPERO. *Cien ejercicios de econometría*, Ediciones Pirámide, España, 1999.

Anexos

Tabla 4.5: Anexo 1: Datos de las series temporales
IRAS, Diarrea y Dengue.

Datos de la serie IRAS desde los años 2007 a 2013							
	2007	2008	2009	2010	2011	2012	2013
Enero	16999	14160	15397	17225	20739	17676	13042
Febrero	19272	15486	14723	19311	17687	14230	18138
Marzo	18558	19197	16269	21595	16852	12854	12801
Abril	18398	14161	20240	19076	16186	23346	19337
Mayo	16300	15078	19819	21147	23702	17512	18476
Junio	16259	24170	24170	21268	17860	21647	19321
Julio	21994	23019	36489	31459	18205	21827	20385
Agosto	22628	14459	22992	19898	17826	18189	12908
Septiembre	12473	18076	18238	15126	15453	19045	16862
Octubre	18247	14888	18749	15151	18698	20214	13328
Noviembre	13509	15003	19594	17985	13756	13595	12674
Diciembre	14312	15117	13184	12325	8761	12421	7702
Datos de la serie Diarrea desde los años 2007 a 2013							
	2007	2008	2009	2010	2011	2012	2013
Enero	2050	1558	2023	1388	2014	3764	1761
Febrero	2673	1844	2121	2384	2198	1754	1844
Marzo	1953	1800	2160	2945	2307	928	1344
Abril	2650	1653	1285	2154	2205	2139	2273
Mayo	3151	1928	2803	4395	2601	3831	3095
Junio	2938	2754	3624	2864	1845	3335	5193

Julio	2255	1887	1118	1424	1546	1633	3067
Agosto	1240	1245	904	1026	1647	1201	1361
Septiembre	1099	1345	787	730	1177	1089	1723
Octubre	1211	993	991	638	1084	1101	979
Noviembre	760	814	1189	988	946	1632	1028
Diciembre	1090	1280	838	989	1264	1960	654
Datos de la serie Dengue desde los años 2007 a 2013							
	2007	2008	2009	2010	2011	2012	2013
Enero	51	21	8	73	19	34	57
Febrero	23	19	2	88	10	17	24
Marzo	9	11	1	212	7	12	34
Abril	20	3	3	112	2	38	71
Mayo	28	4	5	129	15	49	92
Junio	128	7	8	137	23	90	133
Julio	270	8	9	202	63	210	276
Agosto	247	11	85	255	151	283	180
Septiembre	262	19	98	135	141	462	172
Octubre	198	198	188	104	196	278	207
Noviembre	94	94	172	76	79	113	145
Diciembre	56	56	81	19	27	46	24

Tabla 4.6: ANEXO 2: Datos y predicciones de casos de IRAS mensual

Año-Mes	Iras	Predicciones	Int. C	Int. C
2008-01	14160	16999		
2008-02	15486	18705.48		
2008-03	19197	17589.81		
2008-04	14161	17752.07		
2008-05	15078	15358.87		
2008-06	24170	15404.33		
2008-07	23019	22191.03		
2008-08	14459	22985.85		
2008-09	18076	11997.72		
2008-10	14888	18585.58		
2008-11	15003	13426.75		
2008-12	15117	14491.94		
2009-01	15397	15338.91		
2009-02	14723	17088.71		
2009-03	16269	18473.21		
2009-04	20240	15387.06		
2009-05	19819	15697.54		
2009-06	24170	20899.69		
2009-07	36489	23858.99		

Continuación Anexo 2.

Año-Mes	IRAS	Predicciones	Int. C	Int. C
2009-08	22992	19913.41		
2009-09	18238	17293.18		
2009-10	18749	18579.91		
2009-11	19594	16144.73		
2009-12	13184	16982.92		
2010-01	17225	16844.34		
2010-02	19311	17573.11		
2010-03	21595	19573.64		
2010-04	19076	19583.40		
2010-05	21147	18978.81		
2010-06	21268	24137.02		
2010-07	31459	29850.41		
2010-08	19898	21523.40		
2010-09	15126	17573.77		
2010-10	15151	18070.28		
2010-11	17985	16537.41		
2010-12	12325	14456.14		
2011-01	20739	16090.69		
2011-02	17687	17903.05		

Continuación Anexo 2.

Año-Mes	IRAS	Predicciones	Int. C	Int. C
2011-03	16852	19776.87		
2011-04	16186	18329.03		
2011-05	23702	18426.09		
2011-06	17860	22169.43		
2011-07	18205	28760.40		
2011-08	17826	18468.11		
2011-09	15453	14820.12		
2011-10	18698	15491.24		
2011-11	13756	16363.13		
2011-12	8761	12691.38		
2012-01	17676	15796.28		
2012-02	14230	15992.45		
2012-03	12854	16615.39		
2012-04	23346	15346.12		
2012-05	17512	19046.53		
2012-06	21647	18847.71		
2012-07	21827	23534.38		
2012-08	18189	18110.52		
2012-09	19045	14993.48		

Continuación Anexo 2.

Año-Mes	IRAS	Predicciones	Int. C	Int. C
2012-10	20214	17009.18		
2012-11	13595	15941.43		
2012-12	12421	11448.95		
2013-01	13042	17619.99		
2013-02	18138	15564.86		
2013-03	12801	15941.11		
2013-04	19337	18332.10		
2013-05	18476	18534.07		
2013-06	19321	19926.03		
2013-07	20385	22723.67		
2013-08	12908	17819.88		
2013-09	16862	15246.62		
2013-10	13328	16828.12		
2013-11	12674	13448.56		
2013-12	7702	10614.81		
2014-01		13693.83	9431.57	19882.28
2014-02		14600.35	10016.02	21282.94
2014-03		13097.47	8953.32	19159.75
2014-04		16957.22	11555.47	24884.12

Continuación Anexo 2.

Año-Mes	IRAS	Predicciones	In.t C	Int. C
2014-05		16822.63	11431.72	24755.75
2014-06		18016.96	12212.84	26579.49
2014-07		20179.57	13648.41	29836.12
2014-08		14963.02	10100.13	22167.20
2014-09		15259.64	10282.13	22646.74
2014-10		14953.46	10059.89	22227.49
2014-11		12980.36	8720.15	19321.89
2014-12		9456.15	6344.58	14093.75
2015-01		14008.98	9134.01	21485.81
2015-02		14918.44	9702.38	22938.66
2015-03		13367.58	8674.15	20600.57
2015-04		17288.31	11195.69	26696.53
2015-05		17133.59	11075.50	26505.30
2015-06		18332.27	11831.29	28405.36
2015-07		20513.94	13220.25	31831.57
2015-08		15197.76	9781.62	23612.82
2015-09		15486.30	9955.85	24088.89
2015-10		15163.77	9738.46	23611.52
2015-11		13153.21	8439.45	20499.81
2015-12		9575.39	6138.75	14935.97

Tabla 4.11: ANEXO 3: Datos y Predicciones de casos de *Diarrea*

AÑO-MES	Diarrea	Predicciones.	Int. C	Int. C
2008-01	1558	2050		
2008-02	1844	2386.202		
2008-03	1800	1717.324		
2008-04	1653	2715.73		
2008-05	1928	2425.389		
2008-06	2754	2602.173		
2008-07	1887	2323.717		
2008-08	1245	1110.581		
2008-09	1345	1167.541		
2008-10	993	1305.221		
2008-11	814	657.566		
2008-12	1280	1220.413		
2009-01	2023	1597.83		
2009-02	2121	2089.394		
2009-03	2160	1814.368		
2009-04	1285	1812.901		
2009-05	2803	1606.78		
2009-06	3624	3697.735		
2009-07	1118	1866.98		

Continuación Anexo 3.

Año-Mes	Diarrea	Predicciones	Int. C	Int .C
2009-08	904	948.948		
2009-09	787	1310.88		
2009-10	991	757.896		
2009-11	1189	938.202		
2009-12	838	1451.081		
2010-01	1388	1512.616		
2010-02	2384	2026.6		
2010-03	2945	2354		
2010-04	2154	1446.825		
2010-05	4395	3460.533		
2010-06	2864	4113.037		
2010-07	1424	923.005		
2010-08	1026	1137.323		
2010-09	730	745.221		
2010-10	638	980.23		
2010-11	988	947.151		
2010-12	989	856.948		
2011-01	2014	1497.438		
2011-02	2198	2789.096		

Continuación Anexo 3.

Año-Mes	Diarrea	Predicciones.	Int.C	Int. C
2011-03	2307	2596.044		
2011-04	2205	2023.481		
2011-05	2601	4599.553		
2011-06	1845	2117.739		
2011-07	1546	1323.742		
2011-08	1647	1113.888		
2011-09	1177	897.979		
2011-10	1084	736.287		
2011-11	946	1212.579		
2011-12	1264	867.166		
2012-01	3764	2458.758		
2012-02	1754	2753.958		
2012-03	928	1816.755		
2012-04	2139	1544.958		
2012-05	3831	3090.018		
2012-06	3335	2067.423		
2012-07	1633	1991.484		
2012-08	1201	1482.697		
2012-09	1089	1052.734		

Continuación Anexo 3.

Año-Mes	Diarrea	Predicciones.	Int. C	Int. C
2012-10	1101	1103.617		
2012-11	1632	944.812		
2012-12	1960	1688.287		
2013-01	1761	4073.511		
2013-02	1844	1125.036		
2013-03	1344	1205.546		
2013-04	2273	2265.756		
2013-05	3095	3837.481		
2013-06	5193	2976.08		
2013-07	3067	2192.879		
2013-08	1361	1434.486		
2013-09	1723	1059.093		
2013-10	979	1424.642		
2013-11	1028	1337.969		
2013-12	654	1704.703		
2014-01		1060.312	541.54	2076.06
2014-02		1844	862.12	3944.16
2014-03		1344	628.36	2874.70
2014-04		2273	1062.69	4861.76

Continuación Anexo 3.

Año-Mes	Diarrea	Predicciones.	Int. C	Int. C
2014-05		3095	1446.99	6619.94
2014-06		5193	2427.87	11107.39
2014-07		3067	1433.90	6560.05
2014-08		1361	636.30	2911.06
2014-09		1723	805.55	3685.35
2014-10		979	457.71	2094.00
2014-11		1028	480.62	2198.81
2014-12		654	305.76	1398.85
2015-01		1060.312	384.39	2924.76
2015-02		1844	629.21	5404.13
2015-03		1344	458.60	3938.80
2015-04		2273	775.59	6661.37
2015-05		3095	1056.08	9070.38
2015-06		5193	1771.96	15218.88
2015-07		3067	1046.52	8988.32
2015-08		1361	464.40	3988.62
2015-09		1723	587.92	5049.52
2015-10		979	334.06	2869.11
2015-11		1028	350.78	3012.71
2015-12		654	223.16	1916.65

Tabla 4.16: Anexo 4: Datos y Pronósticos para la serie Dengue.

Año-Mes	Dengue	Pronóst.	Int. C	Int. C
2008-01	21	51		
2008-02	19	11.1695		
2008-03	11	7.8524		
2008-04	3	24.3118		
2008-05	4	5.9568		
2008-06	7	26.4072		
2008-07	8	25.4477		
2008-08	11	14.3207		
2008-09	19	21.8759		
2008-10	198	25.5515		
2008-11	94	104.4193		
2008-12	56	63.8479		
2009-01	8	24.8239		
2009-02	2	6.3784		
2009-03	1	1.3878		
2009-04	3	1.1642		
2009-05	5	4.8067		
2009-06	8	16.2618		
2009-07	9	15.6295		

Continuación Anexo 4.

Año-Mes	Dengue	Pronóst.	Int. C	Int. C
2009-08	85	13.216		
2009-09	98	106.2789		
2009-10	188	252.9147		
2009-11	172	90.0086		
2009-12	81	92.8706		
2010-01	73	17.8914		
2010-02	88	26.7315		
2010-03	212	30.9337		
2010-04	112	94.8054		
2010-05	129	92.3398		
2010-06	137	124.9787		
2010-07	202	96.7705		
2010-08	255	423.0102		
2010-09	135	255.0479		
2010-10	104	511.2933		
2010-11	76	75.1195		
2010-12	19	43.8553		
2011-01	19	8.1598		
2011-02	10	10.5614		

Continuación Anexo 4.

Año-Mes	Dengue	Pronóst.	Int. C	Int. C
2011-03	7	11.1501		
2011-04	2	10.0474		
2011-05	15	3.9621		
2011-06	23	21.2679		
2011-07	63	31.1153		
2011-08	151	208.4256		
2011-09	141	118.4621		
2011-10	196	167.8008		
2011-11	79	152.3645		
2011-12	27	29.0936		
2012-01	34	27.2887		
2012-02	17	27.8868		
2012-03	12	24.7063		
2012-04	38	5.7286		
2012-05	49	94.3701		
2012-06	90	61.4258		
2012-07	210	166.7337		
2012-08	283	328.0953		
2012-09	462	186.2219		

Continuación Anexo 4.

Año-Mes	Dengue	Pronóst.	Int. C	Int. C
2012-10	278	389.4936		
2012-11	113	136.2896		
2012-12	46	30.9268		
2013-01	57	45.815		
2013-02	24	25.6231		
2013-03	34	15.2555		
2013-04	71	25.2428		
2013-05	92	159.7411		
2013-06	133	125.8988		
2013-07	276	282.1297		
2013-08	180	432.4285		
2013-09	172	225.2399		
2013-10	207	169.2413		
2013-11	145	85.3639		
2013-12	24	50.2808		
2014-01		31.8832	6.694069599	151.8565006
2014-02		15.4364	2.015864227	118.2041247
2014-03		15.9604	1.582557169	160.9646861
2014-04		42.7921	3.562156073	514.0605191

Continuación Anexo 4.

Año-Mes	Dengue	Pronóst.	Int. C	Int. C
2014-05		57.0025	4.227031611	768.6909476
2014-06		95.5744	6.555117223	1393.486278
2014-07		215.1472	13.98902369	3308.903022
2014-08		207.3628	12.99557611	3308.767359
2014-09		264.8053	16.17747755	4334.53666
2014-10		225.9851	13.56302039	3765.330807
2014-11		120.9323	7.16878368	2040.042663
2014-12		32.0605	1.88417112	545.5312006
2015-01		41.3949	2.087416131	820.8874409
2015-02		18.7679	0.85395652	412.473516
2015-03		22.9298	0.97273168	540.5153304
2015-04		54.3008	2.195162334	1343.21732
2015-05		71.5687	2.79831374	1830.416838
2015-06		111.5297	4.260859924	2919.337541
2015-07		241.3888	9.074225208	6421.320109
2015-08		190.7511	7.090288281	5131.811527
2015-09		210.2414	7.753411965	5700.903785
2015-10		214.3948	7.863124134	5845.655414
2015-11		131.9144	4.819450898	3610.660505
2015-12		27.4823	1.001347908	754.259143