



**UNIVERSIDAD LAICA “ELOY ALFARO” DE MANABÍ
FACULTAD DE CIENCIAS INFORMÁTICAS
CARRERA DE INGENIERÍA EN SISTEMAS**



**DESARROLLO DE UN SISTEMA WEB QUE MUESTRE UN
MODELO PREDICTIVO PARA EL DIAGNÓSTICO DE
PACIENTES CON DIABETES EN EL CANTÓN MANTA.**

**TRABAJO DE TITULACIÓN MODALIDAD PROYECTO
INTEGRADOR, PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERAS EN SISTEMAS**

AUTORES:

**LÁZARO MURILLO JOHANNA LISSETH
LAZO ALAY ANA MARÍA**

DIRECTOR DEL TEMA:

ING. RIVADENEIRA ZAMBRANO FABRICIO JAVIER, Mg.

**MANTA – ECUADOR
Septiembre, 2018**

Certificación

En calidad de docente tutor de la Facultad de Ciencias Informáticas de la Universidad Laica “Eloy Alfaro” de Manabí, certifico:

Haber dirigido y revisado el trabajo de titulación, cumpliendo el total de 60 horas, bajo la modalidad de PROYECTO INTEGRADOR, cuyo tema del proyecto es: “DESARROLLO DE UN SISTEMA WEB QUE MUESTRE UN MODELO PREDICTIVO PARA EL DIAGNÓSTICO DE PACIENTES CON DIABETES EN EL CANTÓN MANTA”, el mismo que ha sido desarrollado de acuerdo a los lineamientos internos de la modalidad en mención y en apego al cumplimiento de los requisitos exigidos por el Reglamento de Régimen Académico, por tal motivo CERTIFICO, que el mencionado proyecto reúne los méritos académicos, científicos y formales, suficientes para ser sometido a la evaluación del tribunal de titulación que designe la autoridad competente.

La autoría del tema desarrollado, corresponde a la Srta. Lázaro Murillo Johanna Liseth y a la Sra. Lazo Alay Ana María, estudiantes de la carrera Ingeniería en Sistemas, periodo académico 2018-2019, quienes se encuentran actas para la sustentación de su trabajo de titulación.

Particular que certifico para los fines consiguientes, salvo disposición de Ley en contrario.

Manta, 06 de septiembre del 2018

Lo certifico,

Ing. Fabricio Rivadeneira Zambrano, Mg.
Docente Facultad de Ciencias Informáticas
Universidad Laica Eloy Alfaro de Manabí

Aprobación del Tribunal de Sustentación

TRABAJO DE TITULACIÓN MODALIDAD PROYECTO INTEGRADOR,
PREVIO A LA OBTENCIÓN DEL TÍTULO DE: INGENIERAS EN SISTEMAS.

“DESARROLLO DE UN SISTEMA WEB QUE MUESTRE UN MODELO
PREDICTIVO PARA EL DIAGNÓSTICO DE PACIENTES CON DIABETES
EN EL CANTÓN MANTA”.

Tribunal examinador que declara **APROBADO** el Grado de **INGENIERA**
EN SISTEMAS, de la Srta. **LÁZARO MURILLO JOHANNA LISSETH** y de la Sra.
LAZO ALAY ANA MARÍA.

Ing. Robert Moreira Centeno

Ing. Oscar González López

Lic. Dolores Muñoz Verduga

Manta, 11 de septiembre del 2018

Declaración Expresa De Autoría

Nosotras, Johanna Lisseth Lázaro Murillo y Ana María Lazo Alay, en calidad de autoras del trabajo de titulación: “DESARROLLO DE UN SISTEMA WEB QUE MUESTRE UN MODELO PREDICTIVO PARA EL DIAGNÓSTICO DE PACIENTES CON DIABETES EN EL CANTÓN MANTA”, autorizamos a la Universidad Laica “Eloy Alfaro” de Manabí, hacer uso parcial o completo del contenido de este trabajo de titulación del cual somos responsables, con fines estrictamente académicos o de investigación.

Los derechos que como autoras nos corresponden, con excepción de la presente autorización, seguirán vigentes a nuestro favor, de conformidad con lo establecido en los artículos 5, 6, 8, 19 y demás artículos pertinentes de la ley de Propiedad Intelectual y su Reglamento.

Asimismo, autorizamos a la Universidad Laica “Eloy Alfaro” de Manabí que realice la digitación y publicación de nuestro trabajo de titulación en el repositorio virtual, en conformidad a lo establecido en el Art. 144 de la Ley Orgánica de Educación Superior.

Manta, 06 de septiembre del 2018

Johanna Lisseth Lázaro Murillo
C.I: 1316864964
Telf.: 0991307577
Email: e1316864964@live.uleam.edu.ec

Ana María Lazo Alay
C.I: 0950582288
Tel.: 0997361733
Email: e0950582288@live.uleam.edu.ec

DEDICATORIA

*“No basta saber, se debe también aplicar.
No es suficiente querer, se debe también hacer”.*

Johann Wolfgang Goethe

A mi madre, por el esfuerzo y entrega constante realizada día tras día para brindarme una buena educación y que no me falte nada en la vida, por estar siempre a mi lado y motivarme a ser mejor. A mi padre, por ser un ejemplo de liderazgo y por sus consejos dados, guiándome para ser una profesional y una persona de buenos valores y principios. A mis hermanos y a mi familia en general por todo el apoyo brindado.

A mis compañeros y amigos, presentes y pasados, quienes sin esperar nada a cambio compartieron sus conocimientos, alegrías y tristezas; a todas aquellas personas que durante estos cinco años estuvieron a mi lado apoyándome, cada momento vivido es simplemente único, cada oportunidad de corregir un error, la oportunidad de que cada mañana pueda empezar de nuevo sin importar la cantidad de errores y faltas cometidas durante el día anterior.

Y finalmente... a mí, por toda la paciencia, constancia y entusiasmo; por no desfallecer aun cuando las ganas de darse por vencida eran tentadoras.

Johanna Lázaro Murillo

DEDICATORIA

A Dios

Por darme la vida y no dejarme desfallecer cada vez que sentía que la vida no tenía sentido poniendo en mí camino a personas que con amor mostraban una luz.

A mis padres

Que aunque han cometido muchos errores hicieron su mejor esfuerzo en darnos educación y en luchar porque no nos faltara lo esencial a mis hermanos y a mí. Gracias a mis padres soy quien soy, con mis defectos y virtudes.

A mi esposo

Quien con su amor y apoyo iluminó mi vida, gracias por tu paciencia y comprensión, la ayuda que me has brindado ha sido sumamente importante, estuviste a mi lado inclusive en los momentos y situaciones más tormentosas, siempre ayudándome. Me has ayudado hasta donde te es posible, incluso más que eso. Muchas gracias, amor. Alfredo Quijije.

A mi hijo

Por ser ese pilar fundamental en mi vida, porque con cada sonrisa que me regalas me das fuerzas para continuar por mas agotada que esté, porque con cada pequeña travesura me enseñas de paciencia y con cada beso y abrazo haces más grande mi corazón, te amo Thiago Alfredo.

Mg. Fabricio Rivadeneira

Más que un ingeniero, un gran amigo, los docentes son un pilar fundamental para el aprendizaje. Agradezco a mi tutor de tesis, quien, con sus conocimientos y su gran trayectoria, nos ha sabido guiar con éxito, gracias por sus consejos, paciencia y comprensión.

Ana María Lazo Alay

AGRADECIMIENTOS

Mi gratitud primordial a mis padres, por ser el apoyo fundamental en mi vida y trayectoria estudiantil, al creer en mí y darme ánimos para no caer ante las adversidades, al darme el digno ejemplo de superación y entrega, gracias a ellos puedo decir que he alcanzado una meta más de muchas que vendrán.

A mis familiares, por la sabiduría que me transmitieron en el desarrollo de mi formación profesional.

A mis compañeros de aula y amigos por ser aquel apoyo incondicional y un motivo de felicidad, en especial a mis compañeras Ana Lazo, Carmen Jama y Betsabet Bastidas, porque juntas hemos crecido en experiencia y conocimiento, además por ser mis amigas en el transcurso de mi formación.

Al Ing. Pedro Villamar, quien de manera incondicional me brindó apoyo emocional y fuerzas durante los primeros años de estudiante.

Al director de este proyecto, el Ing. Fabricio Rivadeneira, por su compromiso, paciencia y su enseñanza; además por las ideas, el apoyo y la confianza otorgada.

Johanna Lázaro Murillo

AGRADECIMIENTOS

Agradezco a Dios en primer lugar por permitirme llegar hasta donde estoy, por poner en mi camino a cada una de las personas que han formado parte de mi vida y carrera estudiantil, porque es gracias a ello que soy la persona que soy.

Gracias a mi familia en general y a mi esposo e hijo quienes son mi motor y ganas de superarme día a día.

Gracias a todos los docentes que he tenido alrededor de toda mi carrera estudiantil, con sus aportes y enseñanzas han puesto un granito de arena para que este trabajo sea un éxito; en especial a los Ing. Fabricio Rivadeneira nuestro tutor de tesis e Ing. Oscar por ayudarnos con la información que requeríamos para que este trabajo sea un éxito.

Gracias a mi compañera y amiga Johanna Lázaro por su paciencia y comprensión en este arduo y largo camino que hemos recorrido para culminar nuestro trabajo de tesis.

Gracias a las hermanas de la Congregación Esclavas del Divino Corazón, quienes con su ejemplo de vida diaria me han devuelto la paz interior que tanto necesitaba y que ha sido primordial para poder culminar este trabajo; gracias también a los compañeros de trabajo de la UE Julio Pierregrosse que sin saber aportan conocimientos nuevos y fomentan mi espíritu innovador.

Ana María Lazo Alay

Contenido

Resumen	XIV
INTRODUCCIÓN.....	1
Problematización	2
Planteamiento del problema:	2
I. Ubicación y contextualización.....	2
II. Génesis del problema	2
III. Estado actual del problema	3
IV. Diagrama causa – efecto del problema:	4
OBJETIVOS.....	5
Objetivo General:	5
Objetivos específicos:	5
JUSTIFICACIÓN	6
CAPITULO I	9
MARCO TEÓRICO DE LA INVESTIGACIÓN.....	9
1.1 Antecedentes de investigaciones relacionadas al tema presentado.....	9
1.2 Definiciones conceptuales	11
1.2.1 Diabetes	11
Diabetes Mellitus tipo 1	11
Diabetes Mellitus tipo 2	11
1.2.2 Series temporales.....	11
1.2.3 Estudio de patrones de datos en las series de tiempos.....	11
Estacionario en su media (horizontal).-	12
Tendencia.-.....	12
Componente cíclico.-	14
Componente estacional.-	14
Componente irregular o aleatorio.-	14
1.2.4 Exploración de patrones de datos con análisis de autocorrelación	15
Funciones de autocorrelación y autocorrelación parcial.....	15
1.2.5 Etapas del pronóstico.....	16
1.3 Metodología Box-Jenkins (ARIMA)	17
1.3.1 Modelos autorregresivos	17
1.3.2 Modelos de promedios móviles.....	19
1.3.3 Modelos de promedio móvil autorregresivos	22

1.3.4	Expresión General de un Modelo ARIMA	23
1.3.5	Estrategia de implementación de la construcción del modelo.....	24
1.3.6	Exploración de información	26
1.4	METODOLOGÍA CRISP-DM	27
1.5	Herramientas de análisis de Datos	29
1.5.1	Minitab.....	29
1.6	Herramientas de Desarrollo.....	29
1.6.1	R Studio	29
1.6.2	Entorno R	30
CAPITULO II		32
MARCO INVESTIGATIVO - DISEÑO METODOLÓGICO.....		32
2.1	Métodos de investigación.....	32
2.1.1	Bibliográfico	32
2.1.2	Analítico	32
2.1.3	Elección de la metodología para el proyecto.....	32
2.2	Descripción y tabulación de los datos.....	34
2.3	Análisis e interpretación de datos.	34
CAPÍTULO III		37
MARCO PROPOSITIVO.....		37
3.1	Comprensión del negocio:	37
3.1.1	Determinar los objetivos del negocio	37
3.1.2	Evaluación de la situación.....	37
3.1.3	Determinar los objetivos del pronóstico.	37
3.1.4	Realizar el plan del proyecto.....	38
3.1.5	Evaluación inicial de herramientas y técnicas.....	38
3.2	Comprensión de los datos:	39
3.2.1	Recolectar los datos iniciales	39
3.2.2	Descripción de los datos	39
3.2.3	Verificar la calidad de los datos	40
3.3	Preparación de los datos.....	40
3.3.1	Limpieza de los datos.....	40
3.4	Modelado.....	40
3.4.1	Escoger la técnica de modelado.....	41
3.4.2	Generar el plan de prueba	41
3.4.3	Pruebas de hipótesis referidas a las distribuciones de frecuencias.....	42
3.4.4	Construcción y evaluación del modelo	44

3.4.5	Resultados.....	48
3.4.6	Aplicativo Web en R.....	49
CAPÍTULO IV.....		55
CONCLUSIONES.....		55
RECOMENDACIONES.....		56
Referencias bibliográficas		57
ANEXOS		58

Índice Tablas

Tabla 1. Patrones de autocorrelación y autocorrelación parcial de los procesos de promedio móvil autorregresivos	22
Tabla 2 Evaluación de la metodología	33
Tabla 3 Cifras diarias de incidencias de diabetes en el IESS durante los años 2013 y 2014.....	34
Tabla 4 Distribución Chi cuadrado, en donde p = probabilidad de encontrar un valor mayor o igual que el chi cuadrado tabulado y v = grados de libertad. (Condor, 2009)	43
Tabla 5 Distribución t (Condor, 2009).....	44

Índice Gráficos e Ilustraciones

Figura 1. Tendencias y componentes cíclicos de una serie de tiempo anual, como el costo de una vivienda. (Hanke & Wichern , 2010)	12
Figura 2. Ejemplo de estacionalidad. (Hanke & Wichern , 2010).....	14
Figura 3. Ejemplo de gráficas de autorrelación simple y parcial. (IBM, 2017).....	15
Figura 4. Coeficientes de autocorrelación y autocorrelación parcial de los modelos AR (1) y AR (2). (Hanke & Wichern , 2010).....	19
Figura 5. Coeficientes de autocorrelación y autocorrelación parcial de los modelos MA (1) y MA (2) (Hanke & Wichern , 2010).....	21

Figura 6. Coeficientes de autocorrelación y autocorrelación parcial de un modelo mixto ARMA(1, 1) (Hanke & Wichern , 2010).....	23
Figura 7. Fases y Tareas del Modelo CRISP-DM (Chapman et al., 2000)	27
Figura 8. Fases del modelo de proceso de la metodología CRISP-DM (Chapman et al., 2000)	28
Figura 9. Interfaz del programa Minitab	29
Figura 10. Interfaz de R Studio	30
Figura 11. Preferencia de Metodología según la empresa KDnuggets ⁹	33
Figura 12. Diagrama de Cajas de los datos obtenidos con los valores de cuartiles y mediana.	35
Figura 13. Diagrama de Cajas de los datos obtenidos con valor de la media de los datos obtenidos.	35
Figura 14. Gráfica de series de tiempo de la diabetes total	41
Figura 15. Autocorrelaciones para los datos de la diabetes total.....	42
Figura 16. Desfases de la función de autocorrelación.	42
Figura 17. Ventana de diálogo del Minitab para los ajustes del modelo ARIMA (0,0,0) (0,1,1)	45
Figura 18. Gráficas de Probabilidad normal y auto correlación de residuos ARIMA (0,0,0) (0,1,1)	45
Figura 19. Ventana de diálogo del Minitab para los ajustes del modelo ARIMA (0,0,0) (1,1,0)	46
Figura 20. Gráficas de Probabilidad normal y auto correlación de residuos ARIMA (0,0,0) (1,1,0)	46
Figura 21. Ventana de diálogo del Minitab para los ajustes del modelo ARIMA (0,0,1) (0,1,0)	47
Figura 22. Gráficas de Probabilidad normal y auto correlación de residuos ARIMA (0,0,1) (0,1,0)	48
Figura 23. Pantalla Principal del aplicativo Web de Modelo de Pronóstico ARIMA	49
Figura 24. Pantalla de Autocorrelación de los datos del aplicativo Web de Modelo de Pronóstico ARIMA	50
Figura 25. Pantalla de la gráfica de Pronósticos de los datos del aplicativo Web de Modelo de Pronóstico ARIMA	50
Figura 26. Pantalla de la Matriz de Pronósticos de los datos del aplicativo Web de Modelo de Pronóstico ARIMA	51
Figura 27. Pantalla de la gráfica de Residuos de los Modelos de Pronóstico ARIMA	51

Figura 28. Declaración de librerías y variables input del aplicativo Web de Modelo de Pronóstico ARIMA.	52
Figura 29. Declaración de las variables output del aplicativo Web de Modelo de Pronóstico ARIMA.	52
Figura 30. Declaración de las variables output del aplicativo Web de Modelo de Pronóstico ARIMA.	53
Ilustración 31. Declaración de las variables output del aplicativo Web de Modelo de Pronóstico ARIMA	53

Resumen

El presente proyecto integrador denominado “DESARROLLO DE UN SISTEMA WEB QUE MUESTRE UN MODELO PREDICTIVO PARA EL DIAGNÓSTICO DE PACIENTES CON DIABETES EN EL CANTÓN MANTA” tiene como objetivo desarrollar un sistema web que muestre los pronósticos de padecimiento de Diabetes Mellitus en el Instituto Ecuatoriano de Seguridad Social (IESS) de la ciudad de Manta; los pronósticos se realizaron utilizando la herramienta Minitab, el desarrollo del proceso con la metodología de Box-Jenkins y para mostrar los resultados en la plataforma web se utilizó el entorno de desarrollo R.

El análisis y comprensión de los datos que se pretende realizar aportará al conocimiento de los posibles pacientes y con esta información ayudará a realizar la prevención, control y tratamiento adecuado.

El estudio se desarrolla a partir de la base de datos proporcionada por el IESS, correspondiente a los años 2013-2014, con un total de 272.238 datos registrados, siendo de estos 14.162 pacientes de la ciudad de Manta.

Palabras claves: Análisis - Arima - Comprensión de datos - Minitab - Pronósticos - R.

Contextualización de la Investigación

INTRODUCCIÓN

La diabetes mellitus se ha convertido en uno de los mayores retos de salud pública contemporánea, afecta la calidad de vida de las personas por las diversas complicaciones que de ella se derivan.

Al convertirse en un problema de salud social, se ha suscitado un enorme interés científico y académico, buscando día a día recopilar y analizar todo tipo de información que permita disminuir este conflicto. En las últimas décadas, a partir de los avances de los equipos de computación y de software adecuados, nuevas metodologías en el análisis y pronósticos de series de tiempo han tenido un crecimiento acelerado y precisamente uno de los campos en los que más se emplea es en la medicina.

El empleo de modelos estadísticos predictivos en las ciencias de la salud ha crecido significativamente en los últimos años. Estos aparecen como un vínculo importante entre la estadística y la práctica médica; son de gran ayuda en la toma de decisiones y permiten la creación de diversos sistemas y herramientas útiles para reducir las incertidumbres, garantizar mejores acciones y establecer eficaces medidas de control para la erradicación de las enfermedades.

Problematización

Planteamiento del problema:

I. Ubicación y contextualización

La diabetes es una enfermedad crónica (de larga duración) que afecta la forma en que el cuerpo convierte los alimentos en energía. Al ser una enfermedad denominada silenciosa, puede pasar mucho tiempo antes de que los síntomas comiencen a aparecer. Esta situación conlleva a la presencia de posibles complicaciones severas (cardiovasculares, oculares, renales, neurológicas y dermatológicas) en pacientes diabéticos que pueden ocasionarles la muerte. (Kasper, Braunwald et al. 2006).

Según el Instituto Nacional de estadísticas y Censos (INEC)¹, en el Ecuador la diabetes es la segunda causa de muerte después de las enfermedades isquémicas del corazón; además, indica que entre los años 2007 – 2016 ha crecido 51% el número de fallecidos por diabetes.

II. Génesis del problema

La mala alimentación de la sociedad, así como la falta de información y de control médico, conlleva a que las personas contraigan enfermedades como la diabetes, sobrepeso y obesidad. Si la diabetes no se controla periódicamente con los medicamentos adecuados puede llevar a complicaciones, siendo esta una gran dificultad para dar atención necesaria a la persona afectada.

¹ Órgano rector de la estadística nacional y el encargado de generar las estadísticas oficiales del Ecuador para la toma de decisiones en la política pública.



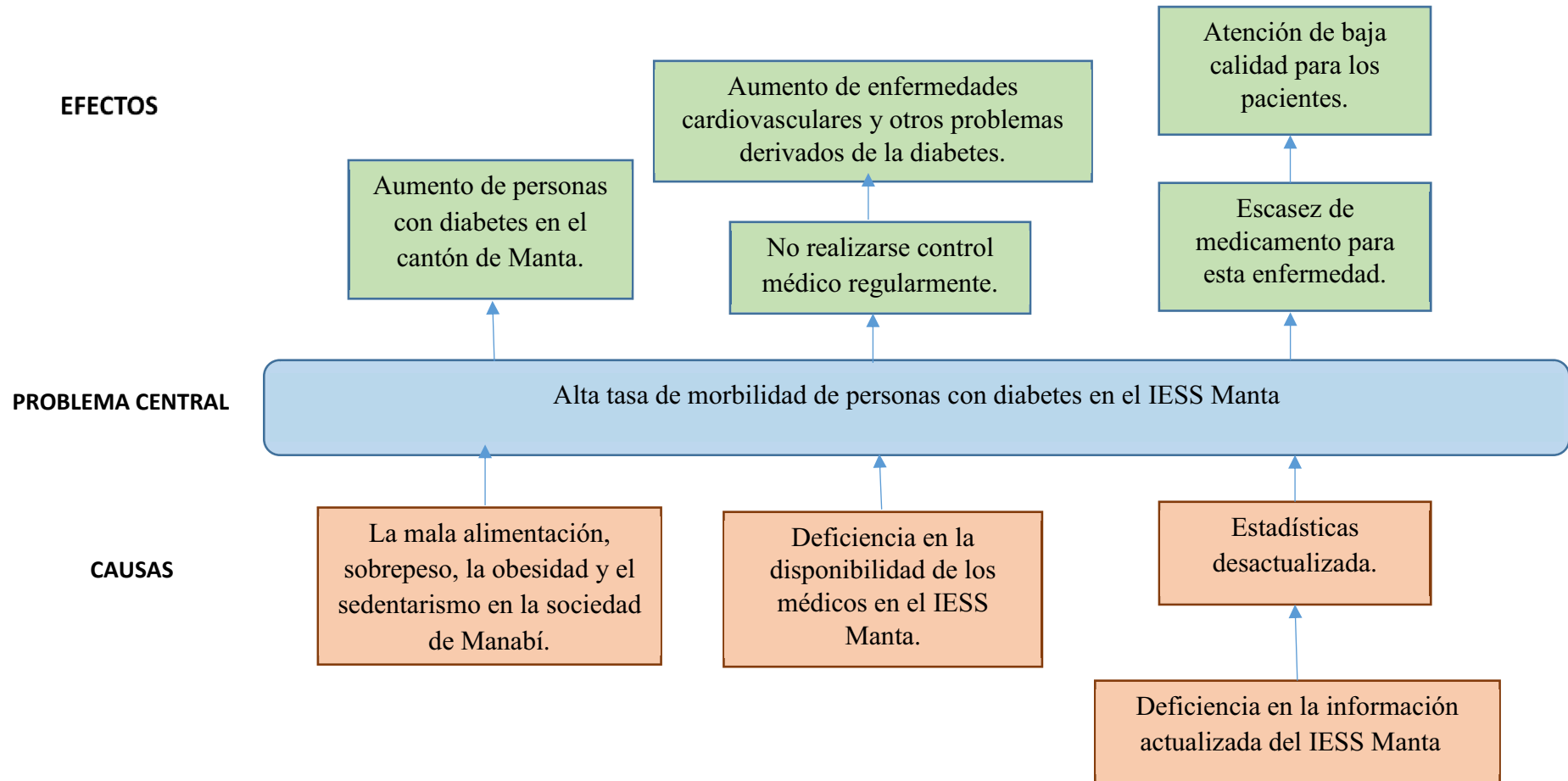
III. Estado actual del problema

Según las estadísticas de los años 2012 – 2013 - 2014 que se presentaron en el programa médico funcional del Hospital IESS Manta, se muestra que la tasa de morbilidad² de esta enfermedad se ha visto en cantidades de 5.341(2012), 1.573(2013), 7.727(2014) personas atendidas en esta institución correspondientemente.

Se observa que no existe una estadística actualizada que permita determinar las cantidades de personas que podrían llegar a requerir de atención médica para esta enfermedad, haciendo que exista escases de medicamentos y disponibilidad médica, además de que se pierda credibilidad en las instituciones debido a que esto incurre en una atención de mala calidad para estos pacientes.

² Cantidad de personas que enferman en un lugar y un período de tiempo determinados en relación con el total de la población.

IV. Diagrama causa – efecto del problema:



OBJETIVOS

Objetivo General:

Desarrollar un sistema web que permita mostrar un modelo de predicción para el diagnóstico de pacientes con diabetes en el cantón Manta.

Objetivos específicos:

- ✓ Analizar los datos para determinar si existe tendencia o estacionalidad.
- ✓ Analizar y validar los diversos modelos de ARIMA en la herramienta Minitab.
- ✓ Aplicar técnicas estadísticas para la comprobación de los modelos en relación a los datos obtenidos.
- ✓ Seleccionar el modelo de mejor ajuste para el pronóstico de pacientes en el IEISS del cantón Manta.
- ✓ Desarrollo de un aplicativo web que muestre los resultados del modelo predictivo.

JUSTIFICACIÓN

La preocupación por la salud es inherente a la condición humana y la ciencia se ha interesado por ella desde su origen; la lucha contra la enfermedad necesita inevitablemente de otras disciplinas que le ayuden a interpretar las diversas dinámicas relativas a la salud.

La Organización Mundial de la Salud (OMS) ha definido la enfermedad como un desequilibrio no sólo físico y psíquico, sino también social, por tanto los métodos de investigación y análisis sociológicos proporcionan las herramientas idóneas para conocer este proceso, que afecta a la calidad de vida de muchos habitantes del cantón Manta.

La correcta identificación de los factores de riesgo asociados a una enfermedad crónica en el Ecuador como lo es la diabetes, puede ayudar a su oportuna prevención. Mediante la creación de niveles de riesgo usando estos factores, las entidades de Salud pueden saber qué población requiere planes específicos de prevención. También al tener un control sobre la población, es posible disminuir la incidencia de la enfermedad o condición crónica y por ende se reducirían también los costos derivados del tratamiento de la enfermedad.

Ser diagnosticado a tiempo ayuda a tomar medidas, tratamiento y cambios en los hábitos de vida que eviten consecuencias nefastas. El 80% de la Diabetes es prevenible mediante la adopción de una dieta saludable y el incremento de la actividad física (Hu, Manson, Stampfer, Colditz, & Willett, 2001).

Por otro lado, diversas técnicas de pronóstico se emplean exitosamente para el diagnóstico, identificación de riesgos y generación de tratamientos de diversas enfermedades.

Estos hechos sustentan el nacimiento de este proyecto, cuyos pronósticos ayudarán a comprender y modelar de una manera más eficiente y precisa la toma de decisiones en la prevención y control de futuros casos de diabetes.

Contenido del trabajo (ideas previas, metodología)

Este proyecto se ha dividido en 4 capítulos; dentro del capítulo I se detallan los antecedentes que se presentan sobre temas similares al proyecto integrador que se está desarrollando, además de un conjunto de conceptos generales, necesarios para comprender el proyecto que se ha planteado, tales como herramientas para análisis, pronósticos de datos y de desarrollo para el sistema web.

Para el capítulo II se detallan los métodos de investigación utilizados para el proyecto. El marco propositivo del capítulo III es la parte más práctica, donde se irá aplicando cada una de las fases de la metodología CRISP-DM al problema que se plantea, que es el pronóstico de las personas con diabetes, y es donde se aplica el modelo seleccionado para el análisis y pronóstico de los datos. En el capítulo IV se encontrarán los resultados obtenidos, las conclusiones y recomendaciones del proyecto.

Finalmente se presenta las referencias bibliográficas, la bibliografía, y anexos.

CAPITULO

I

MARCO TEÓRICO DE LA INVESTIGACIÓN

CAPITULO I

MARCO TEÓRICO DE LA INVESTIGACIÓN

1.1 Antecedentes de investigaciones relacionadas al tema presentado.

En Ecuador existe un gran número de personas que presentan diabetes, así lo demuestra la Encuesta Nacional de Salud realizada en el año 2013, según sus resultados, en el país 414.514 personas mayores de 10 años sufren de diabetes. Según el INEC, se han registrado 4.455 casos de muertes por diabetes mellitus, con una tasa de 23,38 por cada cien mil personas en el 2014. (Mendivil & Sierra, 2011).

La diabetes mellitus es considerada hoy una epidemia mundial por su alta prevalencia, por la disminución en la calidad de vida de quienes la sufren y por ser una de las enfermedades crónicas con alta tasa de fracaso y mala adherencia a su tratamiento. Aun teniendo en cuenta estos puntos, son pocos los estudios relacionados a pronosticar su padecimientos en diversos cantones no solo de la provincia de Manabí, sino también del Ecuador.

Por otra parte, el uso de modelos predictivos en el análisis de información va en un aumento progresivo a nivel internacional, permitiendo aprovechar los patrones de comportamiento encontrados en los datos actuales e históricos para identificar riesgos y oportunidades.

En el año 2016 la Universidad Metropolitana de Ecuador en conjunto con la Universidad de Guayaquil, publicaron el siguiente artículo en la revista electrónica MediSur: **“Modelo estadístico predictivo para el padecimiento de pie diabético en pacientes con diabetes mellitus tipo II”**, se realizó un estudio descriptivo en el período 2010-2013. Como variables, se analizaron los factores de riesgo que influyeron de forma significativa en el

padecimiento de pie diabético. Para la elaboración del modelo se emplearon técnicas multivariadas de regresión logística binaria y árboles de decisión con algoritmo de detector automático de interacciones mediante Chi-cuadrado.

En el año 2014 la Escuela Técnica Superior de Ingenieros de Telecomunicación (E.T.S.I.T), de la Universidad Politécnica de Madrid, publicó la siguiente tesis doctoral en la biblioteca virtual de dicha universidad (Archivo Digital UPM): **“Propuesta de algoritmos de predicción de glucosa en pacientes diabéticos”**, se realizó un análisis comparativo entre la descripción de un modelo de predicción basado en redes neuronales y la descripción de un modelo de predicción basado en técnicas auto-regresivas(ARM). Para evaluar la utilidad clínica de la predicción se ha desarrollado un software denominado GlucoP que implementa el algoritmo de predicción sobre un dispositivo portátil.

En julio del 2016 se publica en la revista tecnológica ESPOL el siguiente tema: **“Proyección de Series de Tiempo para el Consumo de la Energía Eléctrica a Clientes Residenciales en Ecuador”**, teniendo como objetivo principal el presentar una proyección del consumo de electricidad para clientes residenciales del Ecuador para los meses que van desde el año 2013 hasta el año 2022; usando series de tiempo mediante la construcción de modelos ARIMA para el caso univariado y modelos ARIMAX para el caso multivariado. Además, evidenciaron factores que afectan al consumo eléctrico residencial tales como: el precio de electricidad, el ingreso y precio del petróleo.

1.2 Definiciones conceptuales

1.2.1 Diabetes

Es una enfermedad crónica (de larga duración) que afecta la forma en que el cuerpo convierte los alimentos en energía, se caracteriza por la incapacidad del páncreas para producir insulina (hormona indispensable para la vida). (A.D.A.M., 2018)

La Organización Mundial de la Salud (OMS) distingue dos formas principales de diabetes:

Diabetes Mellitus tipo 1 o insulín dependiente, en la que el páncreas no produce insulina, esencial para la vida. Este tipo se desarrolla más a menudo en niños y adolescentes, y al ser incurable, los pacientes requieren de tratamiento (inyecciones de insulina) y seguimiento de por vida para sobrevivir.

Diabetes Mellitus tipo 2, que resulta de la incapacidad del cuerpo para responder adecuadamente a la acción de la insulina que produce el páncreas. Es la más común, alrededor del 90% de casos en el mundo son de este tipo. Se diagnostica mayormente en adultos, pero también están aumentando los casos en adolescentes.

1.2.2 Series temporales

Una serie de tiempo consiste en datos que se recopilan, registran u observan durante incrementos sucesivos de tiempo. (Hanke & Wichern , 2010)

1.2.3 Estudio de patrones de datos en las series de tiempos

De acuerdo con Hanke & Wichern (2010), “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” (p. 82).

Cuando los datos recopilados en el transcurso del tiempo fluctúan alrededor de un nivel o una media constantes, hay un patrón horizontal; se dice que este tipo de series es

estacionario en su media. Cuando los datos no se incrementan ni disminuyen, consistentemente durante un largo periodo tienen un patrón horizontal. Cuando los datos crecen o disminuyen en varios periodos, existe un patrón de tendencia. El componente cíclico es la oscilación alrededor de la tendencia y peculiarmente presenta un ciclo durante varios años. Para más detalle, respecto a los componentes de las series de tiempo se detalla lo siguiente:

Estacionario en su media (horizontal).- Cuando los datos recopilados en el transcurso del tiempo fluctúan alrededor de un nivel o una media constantes.

Tendencia.- Es el componente que representa el crecimiento (o la declinación) subyacente en una serie de tiempo. La tendencia se produce, por ejemplo, como resultado de la inflación, de cambios demográficos persistentes o de cambios tecnológicos e incrementos en la productividad. El componente de la tendencia se denota como T. Para mayor comprensión observe la figura 1.

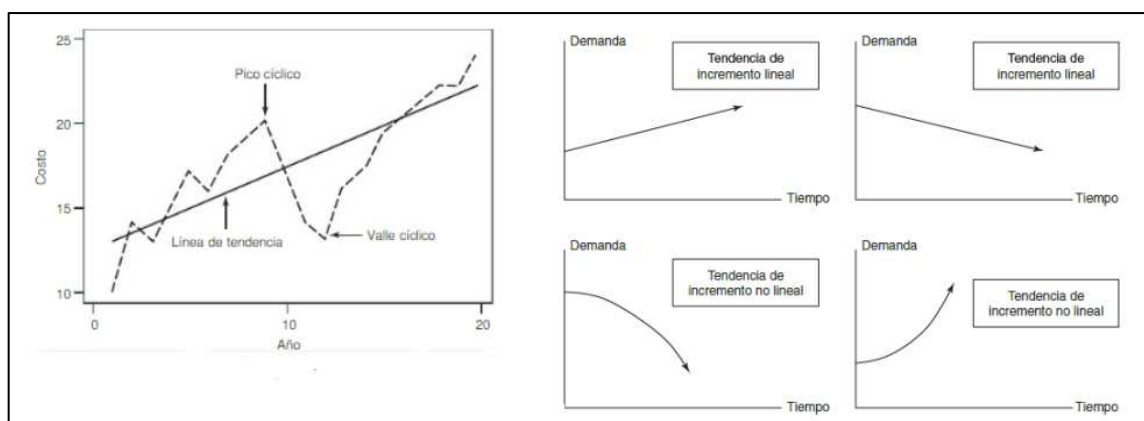


Figura 1. Tendencias y componentes cíclicos de una serie de tiempo anual, como el costo de una vivienda. (Hanke & Wichern, 2010)

A menudo es conveniente ajustar una curva de tendencia a una serie de tiempo por dos razones:

1. Proporciona cierto indicio de la dirección general de la serie observada.

2. Puede eliminarse de la serie original para obtener una imagen más clara de la estacionalidad.

Tendencia lineal. Se utiliza para pronosticar una variable con base en la otra, tanto para pronósticos de series de tiempo como para pronósticos de relaciones causales³; si la tendencia parece ser aproximadamente lineal (es decir, si aumenta o disminuye como una línea recta), entonces se representa por la ecuación:

$$T_t = b_0 + b_1 t \quad (1.1)$$

En la ecuación 1.1, T_t es el valor pronosticado de la tendencia para el tiempo t . El símbolo t representa el tiempo, la variable independiente, y ordinariamente toma los valores 1, 2, 3,... correspondientes a periodos consecutivos. El coeficiente de la pendiente, b_1 , es el incremento o decremento promedio de T para cada incremento de un periodo en el tiempo.

Por ejemplo, cuando los ingresos históricos aumentan o disminuyen a un ritmo constante, se encuentra ante un efecto lineal. Si prevé los ingresos durante los dos próximos trimestres basándose en los ingresos de los cuatro últimos trimestres y si el trazado de multilínea de los ingresos trimestrales anteriores es lineal o casi lineal, el método de tendencia le ofrecerá la previsión más fiable.

Tendencia no lineal. Es un método para encontrar un modelo no lineal para la relación entre la variable dependiente y un conjunto de variables independientes. La tendencia no lineal, puede estimar modelos con relaciones arbitrarias entre las variables independientes y las dependientes.

³ Cuando la ocurrencia del primer evento causa otro. El primer evento es llamado la causa y el segundo evento es llamado efecto. Una correlación entre dos variables no implica causalidad.

Componente cíclico-. Es una serie de fluctuaciones⁴ en forma de onda o ciclos de más de un año de duración. Las condiciones cambiantes de la economía generalmente producen ciclos. La letra C representa el componente cíclico.

Componente estacional-. Los componentes estacionales se encuentran comúnmente en datos trimestrales, mensuales o semanales. La variación estacional se refiere a un patrón de cambio más o menos estable que aparece anualmente y se repite un año tras otro. Los patrones estacionales ocurren por la influencia del tiempo climatológico o por acontecimientos relacionados con el calendario, como las vacaciones escolares y los días feriados nacionales. La figura 2 muestra un ejemplo de estacionalidad anual.

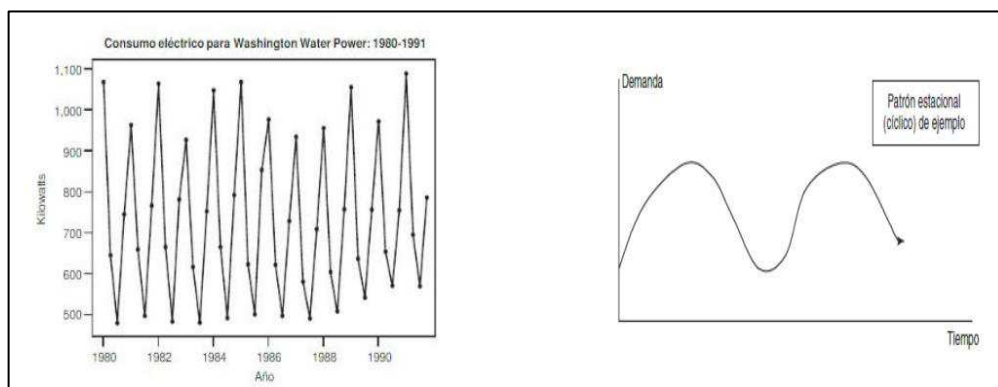


Figura 2. Ejemplo de estacionalidad. (Hanke & Wichern, 2010)

Componente irregular o aleatorio.- Consiste en fluctuaciones impredecibles o aleatorias. Son movimientos erráticos que no siguen un patrón específico y que obedecen a causas diversas. Este componente es prácticamente impredecible. Representan todos los tipos de movimientos de una serie de tiempo que no son tendencia, variaciones estacionales ni actuaciones cíclicas.

⁴ Movimientos oscilatorios alrededor de una tendencia, caracterizados por diferentes fases sucesivas recurrentes, de expansión y contracción, de mayor o menor amplitud, que no se encuentran ceñidas a lapsos fijos y que son susceptibles de medición.

1.2.4 Exploración de patrones de datos con análisis de autocorrelación

En el año 2010, Hanke & Wichern demuestran que “autocorrelación es la correlación que existe entre una variable retrasada uno o más periodos consigo misma” (p. 84).

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.

Funciones de autocorrelación y autocorrelación parcial

La autocorrelación y la autocorrelación parcial son medidas de asociación entre valores de series actuales y pasadas e indican cuáles son los valores de series pasadas más útiles para predecir valores futuros. Con estos datos podrá determinar el orden de los procesos en un modelo ARIMA. (IBM, 2017)

- ✓ *Función de autocorrelación (FAS).* En el retardo k , es la autocorrelación entre los valores de las series que se encuentran a k intervalos de distancia.
- ✓ *Función de autocorrelación parcial (FAP).* En el retardo k , es la autocorrelación entre los valores de las series que se encuentran a k intervalos de distancia, teniendo en cuenta los valores de los intervalos intermedios.

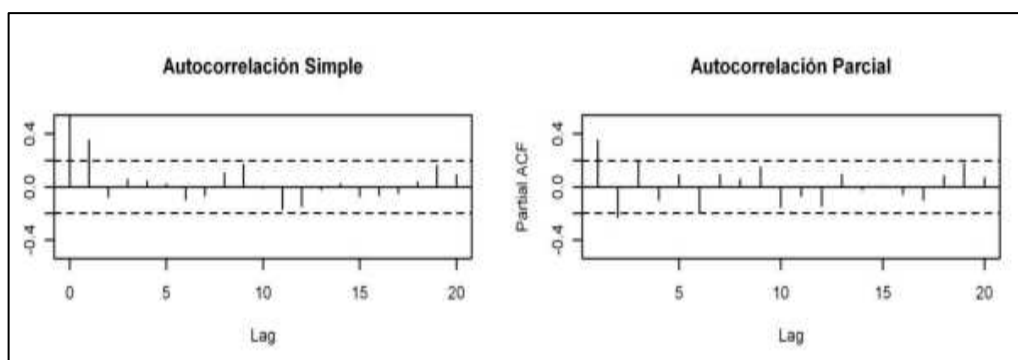


Figura 3. Ejemplo de gráficas de autorrelación simple y parcial. (IBM, 2017)

Una herramienta utilizada con frecuencia para identificar patrones de datos en las series de tiempo es el correlograma o la función de autocorrelación. Un *correlograma* o

función de autocorrelación es una herramienta gráfica para presentar las correlaciones entre diversos retrasos de una serie de tiempo, tal como se observa en la gráfica 3.

1.2.5 Etapas del pronóstico

El reconocimiento de las técnicas de pronósticos funciona con datos generados por eventos históricos. (García, 2014)

1. Formulación del problema y recopilación de datos
2. Manipulación y limpieza de datos
3. Construcción y evaluación del modelo.
4. Implementación del modelo(el pronósticos real)
5. Evaluación del pronóstico.

En la *formulación del problema y la recolección de datos* se tratan como un solo paso porque están íntimamente relacionadas; el problema determina los datos apropiados. Los problemas de recopilación y control de calidad de los datos surgen siempre que sea necesario obtener datos pertinentes para elaborar un pronóstico de negocios.

La *manipulación y limpieza de datos*, a menudo es necesario. Es posible tener demasiados datos o muy pocos, en el proceso para realizar pronósticos. Normalmente se requiere algún esfuerzo para obtener datos en la forma requerida, para usar ciertos procedimientos para pronosticar.

La *construcción y evaluación del modelo*, incluye ajustar los datos recolectados a un modelo de pronóstico que sea adecuado, en términos de minimizar errores en el pronóstico. Cuanto más sencillo sea el modelo, mejor será en términos de la aceptación del proceso de pronósticos por parte de los gerentes que deben tomar las decisiones de la empresa.

La *implementación del modelo*, es la generación del modelo real una vez que se hayan recopilado y depurado los datos apropiados, y se haya seleccionado el modelo de pronóstico

adecuado. Los datos de periodos históricos más recientes se mantienen como respaldo y más tarde se usan para verificar la exactitud del proceso.

En la *evaluación del pronóstico*, implica la comparación de los valores del pronóstico con valores históricos reales. Después de que se termina la implementación del modelo, se realiza los pronósticos para los periodos históricos más recientes, donde se conocen los valores de los datos, pero se mantienen como respaldo de los datos que se analizan. Estos pronósticos se comparan después con los valores históricos conocidos y se analizan cualesquiera errores en el pronóstico. (García, 2014)

1.3 Metodología Box-Jenkins (ARIMA)

Los modelos de promedio móvil integrado autorregresivo (ARIMA), son una clase de modelos lineales capaces de representar tanto series de tiempo estacionarias como no estacionarias. Los modelos se conocen con el nombre genérico de *ARIMA* (AutoRegresive Integrated Moving Average), que deriva de sus tres componentes *AR* (Autoregresivo), *I* (Integrado) y *MA* (Medias Móviles). Expresan un pronóstico como una función de valores previos de la serie de tiempo. Estos modelos utilizan la información de la serie misma para generar los pronósticos. Dependen mucho de los patrones de autocorrelación en los datos. (Fernández, 2016)

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.

1.3.1 Modelos autorregresivos

Los modelos autorregresivos son modelos adecuados para series de tiempo estacionarias, y el coeficiente f_0 está relacionado con el nivel constante de la serie. Los

coeficientes de autocorrelación se aproximan gradualmente hacia cero, mientras que los coeficientes de autocorrelación parcial caen a cero después del primer retraso de tiempo.

Un modelo autorregresivo de orden p tiene la forma

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t \quad (1.2)$$

En donde para la ecuación 1.2

Y_t = Variable de respuesta (o dependiente) en el tiempo t .

$Y_{t-1} + Y_{t-2}, \dots, Y_{t-p}$ = Variable de respuesta en los retrasos de tiempo $t-1, t-2, \dots, t-p$, respectivamente; estas Y desempeñan el papel de variables independientes.

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

ε_t = 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.

El modelo de la ecuación 1.2 tiene un aspecto 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.

La ecuación del modelo AR (1), y del modelo AR (2), se presenta en la figura 4. Las figuras 4a) y b) indican el comportamiento teórico de las funciones de autocorrelación y de autocorrelación parcial, respectivamente, de un modelo AR(1). Los coeficientes de autocorrelación se aproximan gradualmente hacia cero, mientras que los coeficientes de autocorrelación parcial caen a cero después del primer retraso de tiempo. Las figuras 4c) y d) muestran las autocorrelaciones y las autocorrelaciones parciales para un modelo AR (2). Nuevamente los coeficientes de autocorrelación tienden a cero, mientras que los coeficientes de autocorrelación parcial caen a cero después del segundo retraso de tiempo. Este tipo de

patrón generalmente se mantendrá para cualquier modelo AR (p). Debe recordarse que las funciones de autocorrelación muestrales van a diferir de estas funciones teóricas a causa de las variaciones de la muestra.

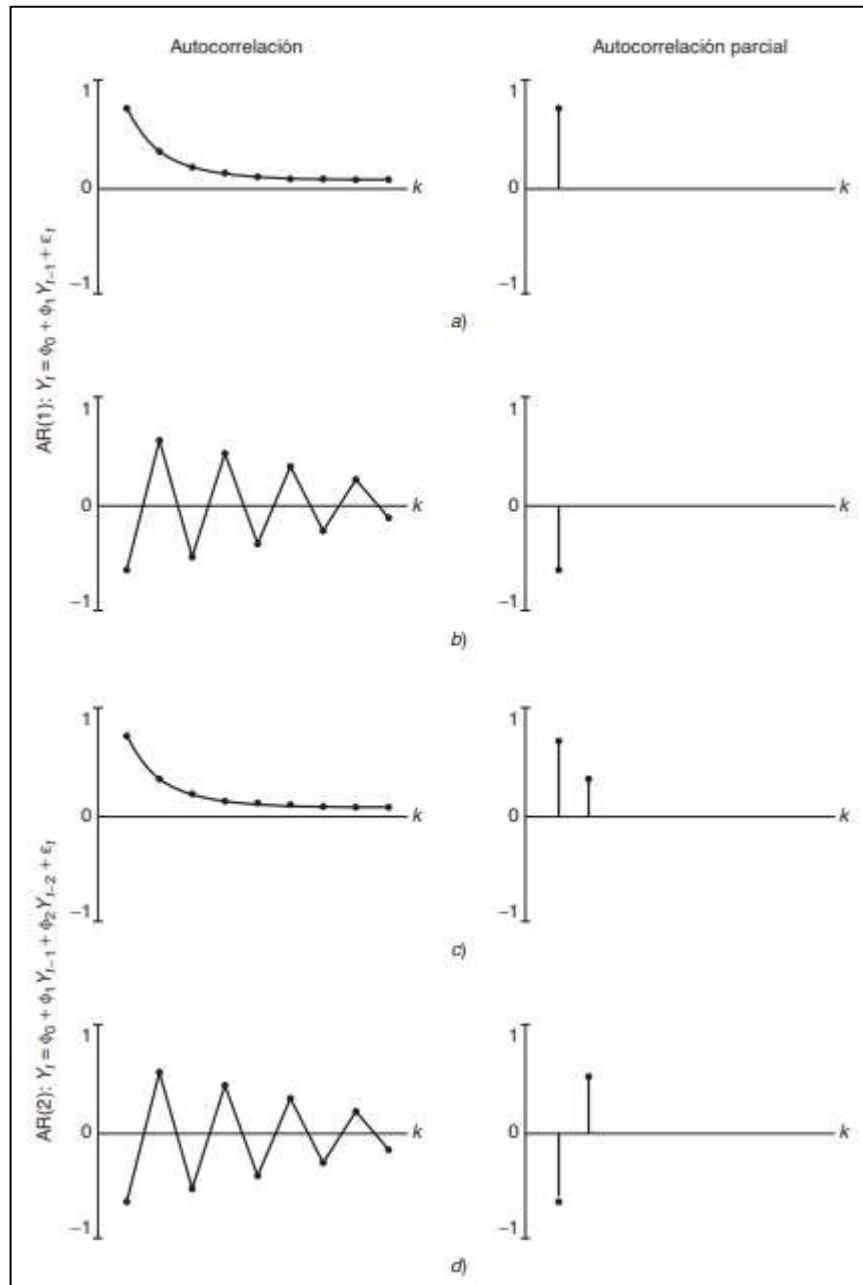


Figura 4. Coeficientes de autocorrelación y autocorrelación parcial de los modelos AR (1) y AR (2). (Hanke & Wichern, 2010)

1.3.2 Modelos de promedios móviles

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

$$Y_t = \mu + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \cdots - \omega_q \varepsilon_{t-q} \quad (1.3)$$

En donde para la ecuación 1.3

Y_t = variable de respuesta (dependiente) en el tiempo t .

μ = promedio constante en el proceso.

$\omega_1, \omega_2, \dots, \omega_q$ = coeficientes que se estimarán.

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.

$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ = errores en periodos anteriores que, para el tiempo t , se incorporan a la respuesta, Y_t

La ecuación 1.3 es similar a la ecuación 1.2, 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 . Este 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. (Hanke & Wichern, 2010)

La figura 5 presenta las ecuaciones de un modelo MA (1) y un modelo MA (2). Los términos de error se pueden agregar en secuencia para obtener un MA (q), donde q es el número de términos de error pasados que se incluirán en el pronóstico de la siguiente observación. Las figuras 5a) y b) también ilustran el comportamiento de los coeficientes teóricos de la autocorrelación y la autocorrelación parcial del modelo MA (1).

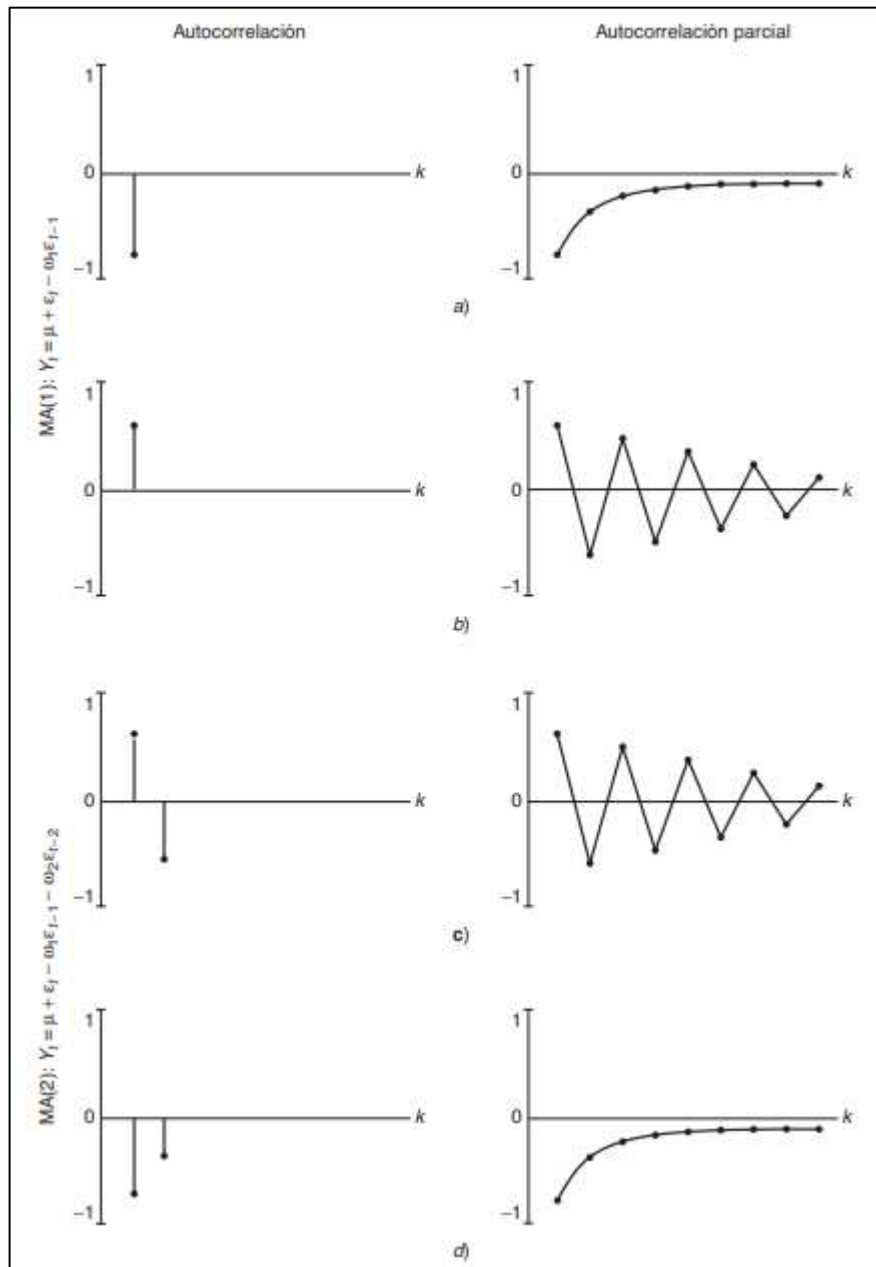


Figura 5. Coeficientes de autocorrelación y autocorrelación parcial de los modelos MA (1) y MA (2) (Hanke & Wichern , 2010)

Nótese que las funciones de autocorrelación y autocorrelación parcial de los modelos AR y MA se comportan de formas muy diferentes. 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 [observe las figuras 5c) y d)].

1.3.3 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. (García, 2014)

Un modelo ARMA (p , q) tiene la forma general de la ecuación 1.4.

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t - \omega_1 \varepsilon_{t-1} - \omega_2 \varepsilon_{t-2} - \dots - \omega_q \varepsilon_{t-q} \quad (1.4)$$

Los modelos ARMA (p , q) pueden describir una gran variedad de comportamientos de las series de tiempo estacionarias⁵. La figura 6 presenta la ecuación de un modelo ARMA (1, 1) y los posibles comportamientos de las autocorrelaciones y autocorrelaciones parciales teóricas.

Los patrones de autocorrelación y autocorrelación parcial de los procesos de promedio móvil autorregresivos se resumen tal como se muestra en la tabla 1.

	Autocorrelaciones	Autocorrelaciones parciales
<i>MA (q)</i>	<i>Terminan después del orden q del proceso</i>	<i>Se desvanecen</i>
<i>AR (p)</i>	<i>Se desvanecen</i>	<i>Terminan después del orden p del proceso</i>
<i>ARMA (p, q)</i>	<i>Se desvanecen</i>	<i>Se desvanecen</i>

Tabla 1. Patrones de autocorrelación y autocorrelación parcial de los procesos de promedio móvil autorregresivos

⁵ Observe 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 .

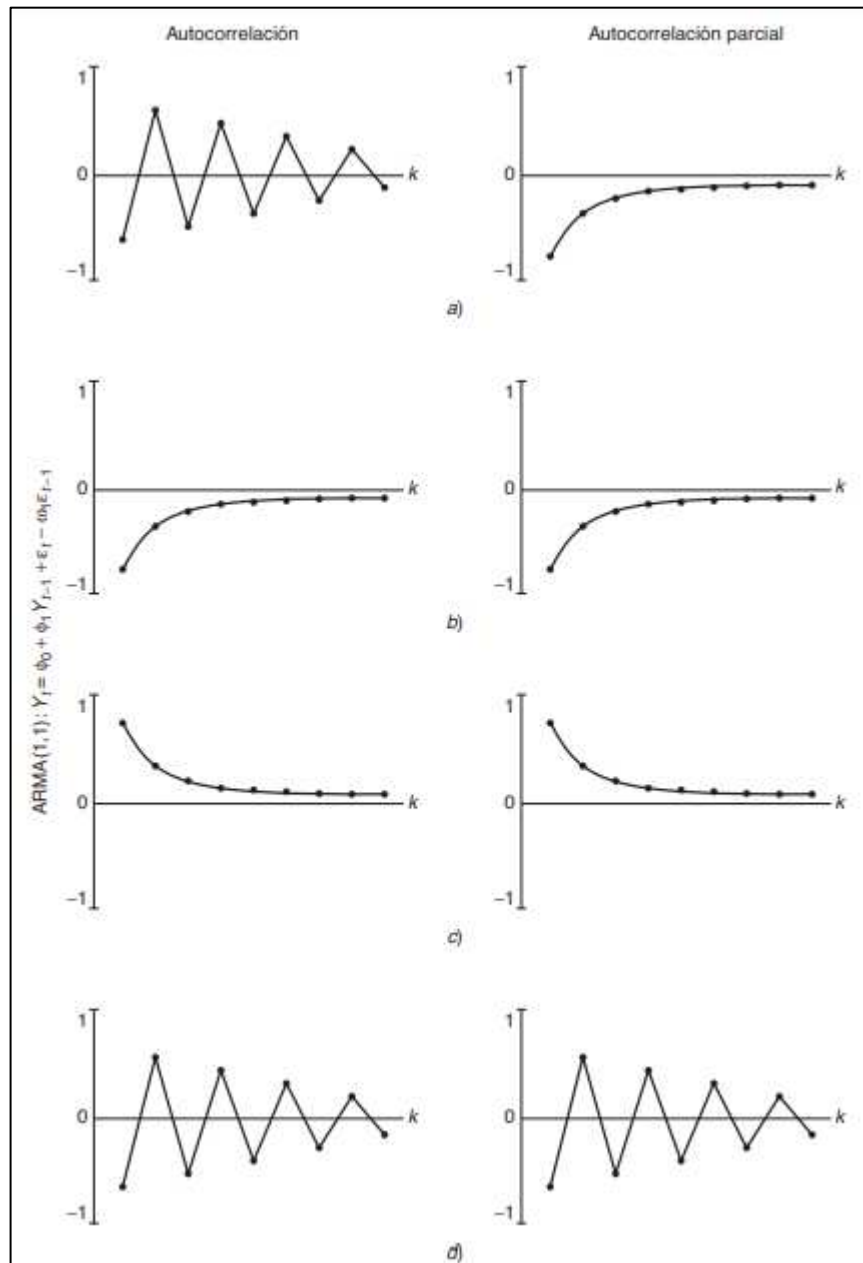


Figura 6. Coeficientes de autocorrelación y autocorrelación parcial de un modelo mixto ARMA(1, 1) (Hanke & Wichern, 2010)

1.3.4 Expresión General de un Modelo ARIMA

Los modelos ARIMA se construyen a partir de los modelos ARMA, pero considerando que la serie en estudio para que sea estacionaria en media tendrá que diferenciarse una serie de veces. Un modelo ARIMA (p,d,q) es un modelo ARMA (p,q) sobre la serie *diferenciada* *d* veces. Es decir, su expresión algebraica será tal como se muestra en la ecuación 1.5

$$Y_t(d) = C + \underbrace{\phi_1 \cdot Y_{t-1}(d) + \dots + \phi_p \cdot Y_{t-p}(d)}_{\text{Comp. Autorregresiva}} + \underbrace{\theta_1 \cdot \varepsilon_{t-1}(d) + \dots + \theta_q \cdot \varepsilon_{t-q}(d)}_{\text{Comp. De Media Móvil}} + \varepsilon_{t-q}(d) \quad (1.5)$$

En donde $Y_t^{(d)}$ es la serie de las diferencias de orden d y $\varepsilon_t(d)$ es la serie de los errores que se cometen en la serie anterior. Habitualmente el orden de diferenciación d , entero, oscila entre 0 y 2.

1.3.5 Estrategia de implementación de la construcción del modelo

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, se está en condiciones de utilizar el modelo ajustado para pronosticar. (Hanke & Wichern, 2010). Los pasos de la estrategia para la construcción del modelo son los siguientes:

Paso 1: Identificación del modelo.- 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. Se recomienda una serie de tiempo no estacionaria si la serie parece crecer o declinar en el tiempo y las autocorrelaciones de la muestra no se desvanecen rápidamente.

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

Paso 2: Estimación del modelo.- 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.

Paso 3: Verificación del modelo.- 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.

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-Box⁶ (LBQ). Esta prueba considera los tamaños de las autocorrelaciones residuales como un grupo. El estadístico de prueba Q es la ecuación 1.6

$$Q = n(n + 2) \sum_{k=1}^m \frac{r_k^2(e)}{n - k} \quad (1.6)$$

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

Para lo cual las variables de la ecuación 1.6 son los siguientes:

$r_k(e)$ = autocorrelación residual para retraso k.

n = número de residuos

k = retraso de tiempo.

⁶ Tipo de prueba estadística para comprobar si una serie de observaciones en un período de tiempo específico son aleatorias e independientes.

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 $p < .05$), el modelo se considera inadecuado. El analista debe considerar un nuevo modelo o uno modificado y continuar el análisis hasta encontrar un modelo satisfactorio.

Paso 4: Elaboración de pronósticos con el modelo. - 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. Los programas de computadora que ajustan modelos ARIMA generan pronósticos e intervalos de predicción a requerimiento del analista.

Conforme más datos están disponibles, se puede usar el mismo modelo ARIMA para generar pronósticos modificados de otro origen de tiempo. Si el patrón de la serie parece estar cambiando con el tiempo, los datos nuevos se pueden usar para recalcular los parámetros del modelo o, si es necesario, desarrollar un modelo completamente nuevo.

1.3.6 Exploración de información

Larose (2005) define el término *explotación de información (Data Mining)* como el proceso de descubrir nuevas correlaciones, patrones y tendencias utilizando grandes cantidades de datos almacenados en repositorios, usando tecnologías de reconocimiento de patrones así como herramientas matemáticas y de estadísticas. Existen actualmente varias metodologías para proyectos de explotación de información, entre ellas podemos nombrar CRISP-DM, y SEMMA como las más reconocidas y algunas otras que no abarcan la totalidad de los proyectos sino que se enfocan en ciertos procesos de los mismos.

1.4 METODOLOGÍA CRISP-DM

La metodología CRISP-DM (Chapman et al., 2000) se encuentra definida en base a un modelo jerárquico de procesos. Esta metodología define un ciclo de vida de los proyectos de explotación de información que define las principales fases de un proyecto de este tipo. Estas fases son: Entendimiento de Negocios, Entendimiento de los Datos, Preparación de los Datos, Modelado, Evaluación y Despliegue. La figura 7 muestra las fases de esta metodología y sus correspondientes tareas, y posteriormente se describirán con mayor detalle.

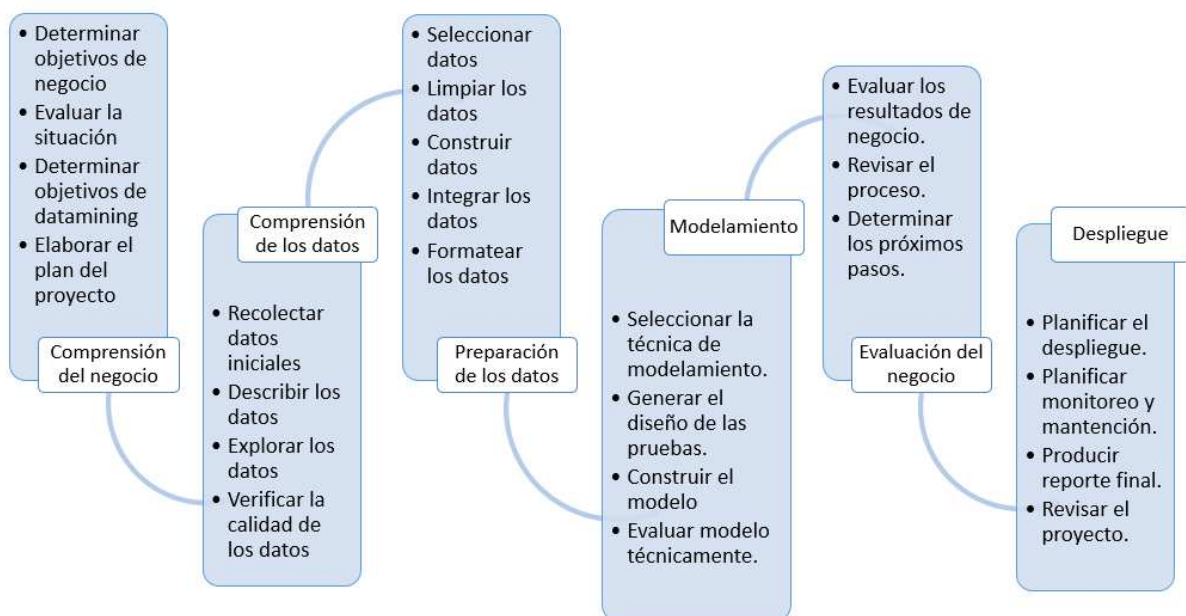


Figura 7. Fases y Tareas del Modelo CRISP-DM (Chapman et al., 2000)

A continuación se presenta el concepto de cada una de las fases identificadas por CRISP-DM. La fase de *Comprensión del Negocio* es la inicial, se centra en la comprensión de los objetivos del proyecto y los requisitos desde una perspectiva de negocio, a continuación, convertir este conocimiento en una definición de problema de minería de datos y un plan preliminar diseñado para lograr los objetivos. La *comprensión de los datos* comienza con una recopilación inicial de datos y prosigue con las actividades para familiarizarse con los datos, identificar problemas de calidad de los datos, descubrir los primeros datos o detectar subconjuntos interesantes para formar hipótesis para Información oculta. La fase de

preparación de datos abarca todas las actividades para construir el conjunto de datos final a partir de los datos iniciales sin procesar.

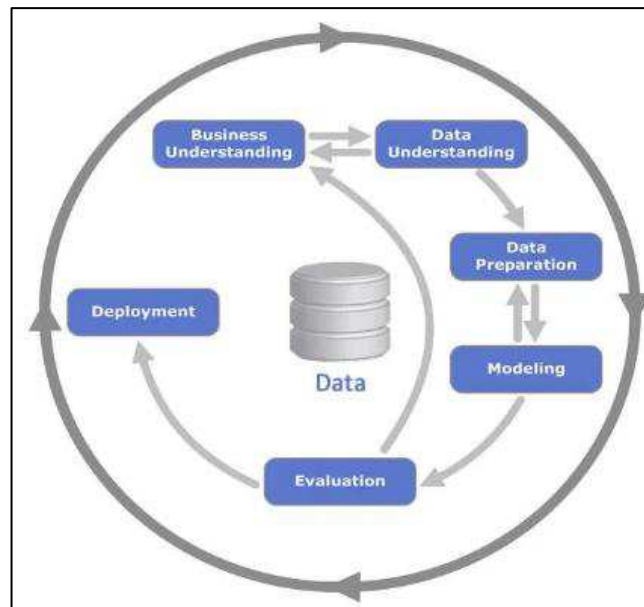


Figura 8. Fases del modelo de proceso de la metodología CRISP-DM (Chapman et al., 2000)

En la fase del *modelado* se seleccionan y aplican diversas técnicas de modelado y se calibran sus parámetros a valores óptimos. Para la etapa de *evaluación el modelo* o modelos obtenidos se evalúan más a fondo y se revisan los pasos ejecutados para construir el modelo para asegurarse de que alcanza adecuadamente los objetivos de negocio. Por último, la fase de *despliegue* nos indica que la creación del modelo generalmente no es el final del proyecto. Incluso si el propósito del modelo es aumentar el conocimiento de los datos, el conocimiento adquirido deberá ser organizado y presentado de manera que el cliente pueda utilizarlo. La figura 8 presenta las fases del modelo de proceso de la metodología CRISP-DM antes descrita.

1.5 Herramientas de análisis de Datos

1.5.1 Minitab

El Programa MINITAB es un programa estadístico muy potente, versátil y de fácil uso, cuenta con una interfaz muy intuitiva, como se puede observar en la figura 3, que proporciona un amplio rango de aplicaciones estadísticas, capacidad de gráfico. Incluye análisis de datos exploratorios, cálculos básicos, regresión, análisis de varianza, tamaño de muestra, análisis multivariante, distribuciones no-paramétrica, series de tiempo, tabulación cruzada y simulación. (Condor, 2009)

En la figura 9 se puede observar la interfaz de la herramienta estadística minitab.

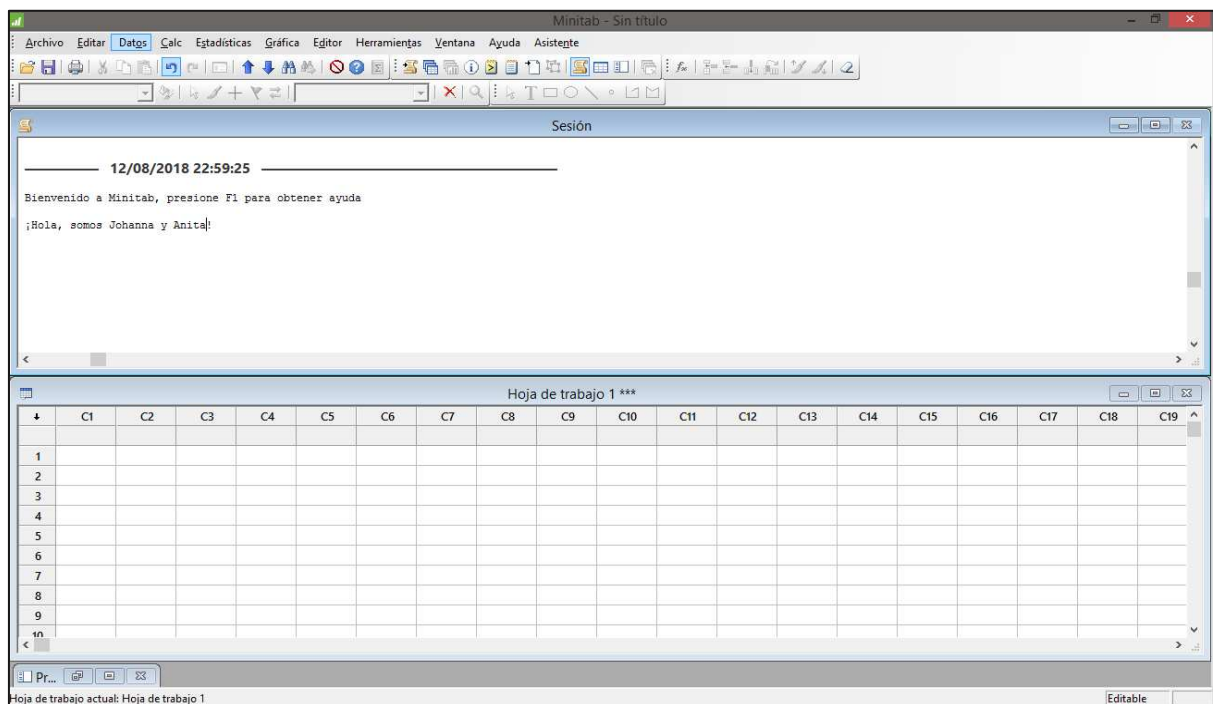


Figura 9. Interfaz del programa Minitab

1.6 Herramientas de Desarrollo

1.6.1 R Studio

RStudio es un entorno de desarrollo integrado (IDE) para el lenguaje de programación R, dedicado a la computación estadística y gráficos. Incluye una consola, editor de sintaxis que apoya la ejecución de código, así como herramientas para el trazado, la depuración y la

gestión del espacio de trabajo. RStudio tiene la misión de proporcionar el entorno informático estadístico R. (Data Carpentry, 2018)

1.6.2 Entorno R

R es un entorno y un lenguaje de programación (véase su interfaz en la figura 10), enfocado en el análisis estadístico de los más utilizados en el campo de la minería de datos que pueden aplicarse a gran variedad de disciplinas. Este lenguaje forma parte de GNU⁷ y se distribuye bajo licencia GNU GLP⁸ (Licencia Pública General de GNU). (Data Carpentry, 2018)

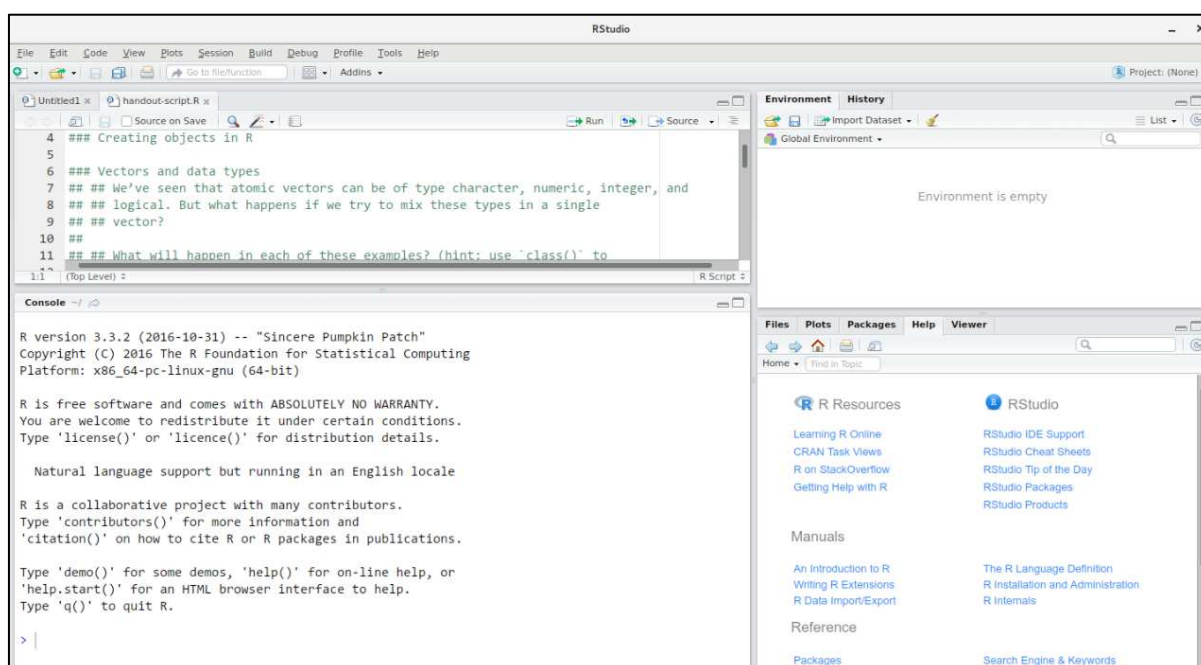


Figura 10. Interfaz de R Studio

⁷ Proyecto colaborativo de software libre con el objetivo de crear un sistema operativo completamente libre: el sistema GNU.

⁸ Licencia de derecho de autor ampliamente usada en el mundo del software libre y código abierto.

CAPITULO

II

MARCO INVESTIGATIVO

(DISEÑO METODOLÓGICO)

CAPITULO II

MARCO INVESTIGATIVO - DISEÑO METODOLÓGICO

2.1 Métodos de investigación

2.1.1 Bibliográfico

La investigación bibliográfica representa una excelente introducción a todos los tipos de investigación, además de que constituye una necesaria primera etapa de todas ellas, puesto que esta proporciona el conocimiento de las investigaciones ya existentes como teorías, hipótesis, experimentos, resultados, instrumentos y técnicas usadas acerca del tema o problema que el investigador se propone investigar o resolver.

2.1.2 Analítico

El Método analítico es aquel método de investigación que consiste en la desmembración de un todo, descomponiéndolo en sus partes o elementos para observar las causas, la naturaleza y los efectos. El análisis es la observación y examen de un hecho particular. Es necesario conocer la naturaleza del fenómeno y objeto que se estudia para comprender su esencia. Este método nos permite conocer más del objeto de estudio, con lo cual se puede: explicar, hacer analogías y comprender mejor su comportamiento y establecer nuevas teorías.

2.1.3 Elección de la metodología para el proyecto

Para elegir la metodología a implementar se realizó un análisis sobre la base de características deseadas, luego de lo cual se elige implementar la metodología que satisface de forma completa los criterios de selección. Se desea una metodología que sea de amplio uso entre profesionales por lo cual se recurre a un análisis de kdnuggets⁹ en el que se puede ver la preferencia de 200 usuarios en el uso de una metodología de minería de datos. Como se observa en la figura 11, en primer lugar de preferencia se encuentra la

⁹ <http://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html>

metodología CRISP-DM con un 43%, en segundo lugar SEMMA con un 8.5% y en tercer lugar KDD con 7.5%.

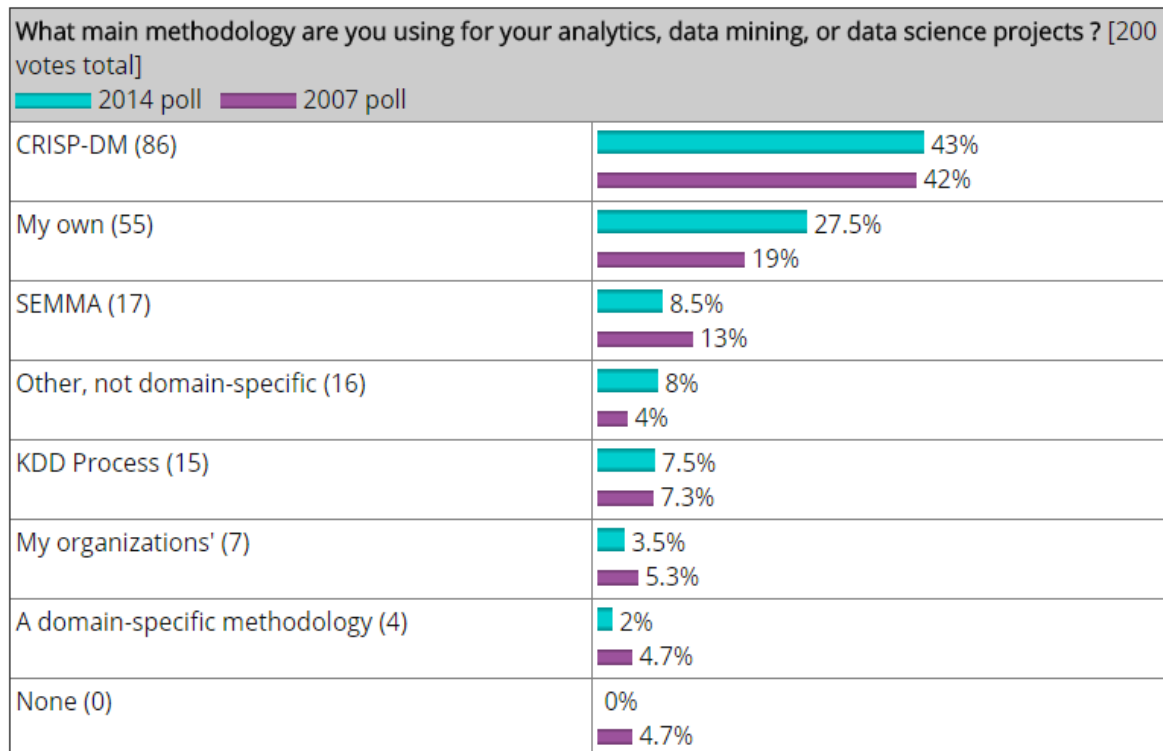


Figura 11. Preferencia de Metodología según la empresa KDnuggets⁹

Además de ser una metodología predominante en su uso, también es necesario ver otros aspectos, los cuales se indican en la tabla 2.

Criterios/Metodologías	CRISP-DM	SEMMA	KDD
Metodología estructurada	✓	✓	✓
Metodología Independiente	✓	X	✓
Ampliamente usada.	✓	X	X
Mejora la calidad de resultados en proyectos de Data Mining.	✓	✓	✓
Herramientas y técnicas independientes.	✓	✓	✓
Finalidad diversa (ejemplo: ampliamente estable en la resolución de problemas variados).	✓	✓	✓
Fácil de implementar.	✓	✓	✓

Tabla 2 Evaluación de la metodología

2.2 Descripción y tabulación de los datos

La tabla 3 corresponde a los 56 promedios diarios del índice de diabetes durante los meses de diciembre y enero de los años 2013-2014 correspondientes.

Tiempo	Datos Originales	Tiempo	Datos originales	Tiempo	Datos originales	Tiempo	Datos originales
1	2	15	48	29	44	43	6
2	1	16	52	30	6	44	1
3	39	17	40	31	52	45	54
4	20	18	44	32	44	46	44
5	31	19	50	33	47	47	37
6	31	20	1	34	47	48	50
7	41	21	53	35	58	49	46
8	3	22	35	36	5	50	5
9	41	23	44	37	3	51	45
10	38	24	62	38	38	52	40
11	61	25	7	39	46	53	37
12	51	26	1	40	46	54	47
13	39	27	15	41	49	55	46
14	5	28	36	42	48		

Tabla 3 Cifras diarias de incidencias de diabetes en el IESS durante los años 2013 y 2014

2.3 Análisis e interpretación de datos.

En la figura 12 se aprecia el diagrama de cajas del cual se puede interpretar que la parte baja es mayor, esto quiere decir que el número de pacientes comprendidas entre el 25% y el 50% de la población está más dispersa que entre el 50% y el 75%, el rango intercuartilico $Q3-Q1$ es igual a 38; es decir el 50% de la población está comprendida en un número de 38

pacientes, la mediana con un valor de 40.5 pacientes nos indica que los datos utilizados siguen una distribución asimétrica.

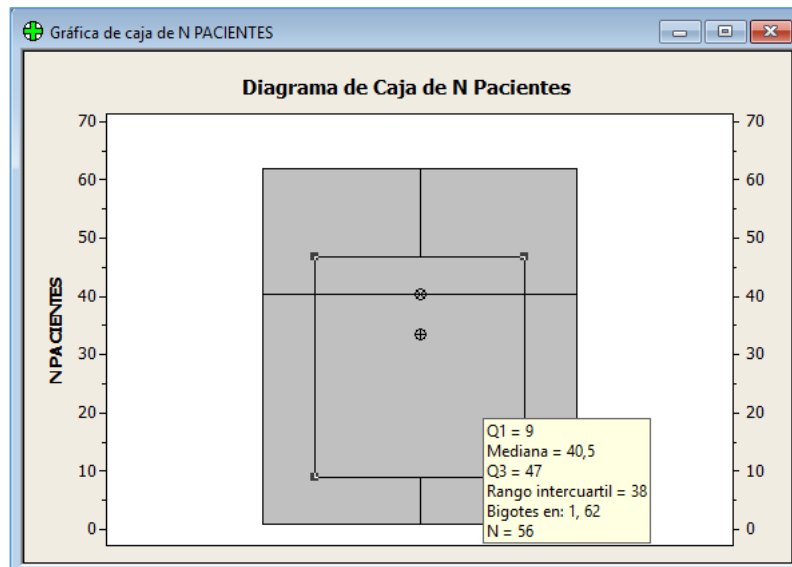


Figura 12. Diagrama de Cajas de los datos obtenidos con los valores de cuartiles y mediana.

En la figura 13 se logra apreciar el valor medio del número de pacientes que nos da un valor de 33, es decir, en promedio se atienden a 33 pacientes con diabetes a diario.

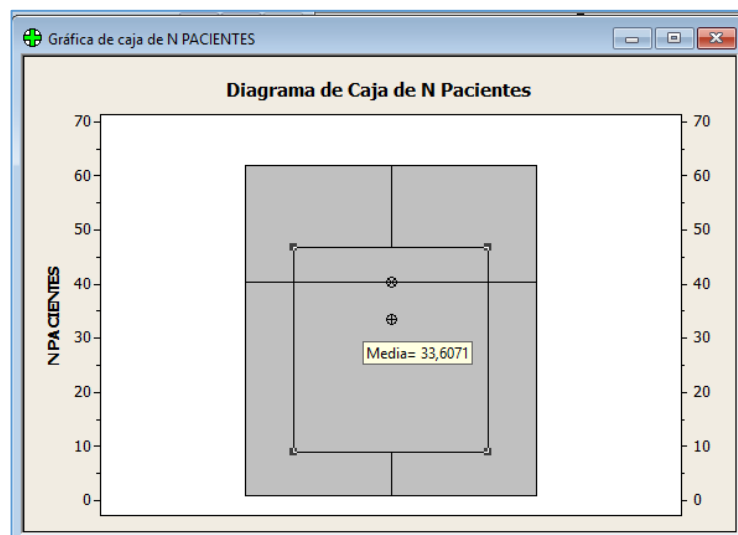


Figura 13. Diagrama de Cajas de los datos obtenidos con valor de la media de los datos obtenidos.

CAPITULO

III

MARCO PROPOSITIVO (ELABORACIÓN DE LA PROPUESTA)

CAPÍTULO III

MARCO PROPOSITIVO

3.1 Comprensión del negocio:

A continuación se irá detallando cada una de las tareas planteadas en esta primera fase.

3.1.1 Determinar los objetivos del negocio

El objetivo de lo que se va a aplicar en este proyecto es el de hacer pronósticos lo más confiables posible, para el estudio de los futuros casos de personas con padecimientos de diabetes, con la finalidad de reducir la tasa de morbilidad en los habitantes del cantón Manta.

3.1.2 Evaluación de la situación

Se cuenta con una base de datos en Excel con información detallada de los pacientes que han sido registrados en el Instituto Ecuatoriano de Seguridad Social durante los años 2013 – 2014, con un total de 272.238 datos registrados, siendo de estos 14.162 pacientes de la ciudad de Manta.

Esta información incluye edad, género, identificación y cantón; pero el dato en que se centrará el pronóstico será en el diagnóstico del paciente.

3.1.3 Determinar los objetivos del pronóstico.

Entre los objetivos tenemos los siguientes:

- ✓ Formulación del problema y recopilación de datos
- ✓ Manipulación y limpieza de datos
- ✓ Construcción y evaluación del modelo.
- ✓ Implementación del modelo(el pronósticos real)
- ✓ Evaluación del pronóstico.

3.1.4 Realizar el plan del proyecto.

El proyecto se dividirá en las siguientes etapas para facilitar su organización y estimar el tiempo de realización del mismo:

Etapa 1:	Análisis de la estructura de los datos y la información de la base de datos. Tiempo estimado: 2 semanas
Etapa 2:	Ejecución de consultas para tener muestras representativas de los datos. Tiempo estimado: 1 semana
Etapa 3:	Preparación de los datos s (selección, limpieza, conversión y formateo, si fuera necesario) para facilitar la minería de datos sobre ellos. Tiempo estimado: 3 semanas.
Etapa 4:	Elección de las técnicas de modelado y ejecución de las mismas sobre los datos. Tiempo estimado: 1 semana.
Etapa 5:	Análisis de los resultados obtenidos en la etapa anterior, si es necesario repetir la etapa 4. Tiempo estimado: 1 semana
Etapa 6:	Producción de informes con los resultados obtenidos en función de los objetivos de negocio y los criterios de éxito establecidos. Tiempo estimado: 1 semana
Etapa 7:	Presentación de los resultados finales. Tiempo estimado: 1 semana.

3.1.5 Evaluación inicial de herramientas y técnicas

La herramienta a utilizar para llevar a cabo este proyecto es Minitab y posteriormente los resultados serán mostrados en una página web desarrollada en R Studio.

En cuanto a las técnicas y modelos predictivos que se van a emplear tenemos las series temporales, aplicando para el desarrollo del proceso, la metodología de Box-Jenkins con un modelo autorregresivo integrado de promedio móvil (ARIMA).

3.2 Comprensión de los datos:

En esta segunda fase de la metodología CRISP-DM se realiza la recolección inicial de los datos para poder establecer nuestro primer contacto con el problema, familiarizarnos con los datos y averiguar su calidad, así como identificar las relaciones más evidentes para formular las primeras hipótesis.

3.2.1 Recolectar los datos iniciales

Los datos utilizados en este proyecto son referentes a los pacientes del IESS de la provincia de Manabí; se encuentran en un archivo csv, en donde se incluye la siguiente información:

- ✓ Cantón al que pertenece
- ✓ Identificación del beneficiario
- ✓ Edad
- ✓ Género
- ✓ Nombre de la Unidad Médica
- ✓ Diagnóstico

3.2.2 Descripción de los datos

Como se ha mencionado con anterioridad, el estudio se desarrolla a partir de la base de datos proporcionada por el IESS, correspondientes a los años 2013-2014, con un total de 272.238 datos registrados.

Variables	Tipo	Descripción
Diagnóstico	Alfanumérico	Tipo de enfermedad del paciente. Este campo es un código que identifica cada enfermedad.

3.2.3 Verificar la calidad de los datos

Los datos se obtuvieron por parte del Ing. Oscar Armando Gonzales López, técnico informático de la Subdirección de Salud del Instituto Ecuatoriano de Seguridad Social (IESS) de Manabí, durante el periodo septiembre del 2011 a Julio del 2014.

3.3 Preparación de los datos

En esta fase de la metodología se trata de preparar los datos para adecuarlos a las técnicas de pronósticos que se van a emplear sobre ellos. Esto implica seleccionar el subconjunto de datos que se va a utilizar, limpiarlos para mejorar su calidad y darles el formato requerido por la herramienta de modelado.

3.3.1 Limpieza de los datos

La base de datos con la que se cuenta para el proyecto contiene toda la información necesaria para poder cumplir los objetivos planteados y no cuenta con campos que tengan valores en 0, o valores negativos.

Contando con una base de datos de 272.238 datos registrados, se realizó el filtrado de datos con la herramienta R, siendo 14.162 pacientes de la ciudad de Manta y se trabajará con el campo “DIAGNÓSTICO”.

3.4 Modelado

En esta fase de la metodología se escogerá la técnica (o técnicas) más apropiadas para los objetivos marcados del pronóstico. A continuación y una vez realizado un plan de prueba para los modelos escogidos, se procederá a aplicar dichas técnicas sobre los datos para generar el modelo y por último se tendrá que evaluar si dicho modelo ha cumplido los criterios de éxito o no.

3.4.1 Escoger la técnica de modelado

Se escogerá la técnica que mejor se adapte a los objetivos propuestos, puesto que los problemas que se desean resolver son problemas de pronóstico, se pasará a comprobar si no se cuenta con datos aleatorios y verificar la aceptación o rechazo la hipótesis nula.

3.4.2 Generar el plan de prueba

Se puede observar en la figura 14 que la serie de tiempo de la diabetes total durante los meses de diciembre y enero de los años 2013 y 2014 correspondientemente, parece variar alrededor de un nivel fijo de 40 (véase la figura 14), aparentemente se presenta una serie estacionaria.

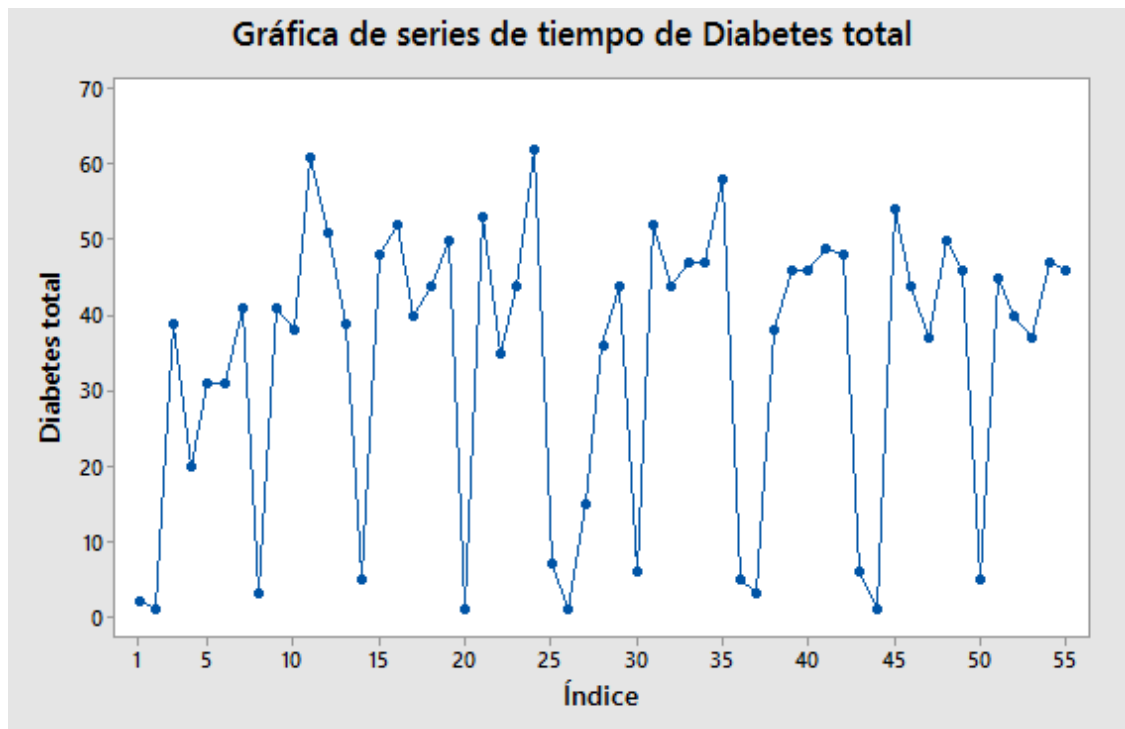


Figura 14. Gráfica de series de tiempo de la diabetes total

A continuación, para confirmar si es estacionario se procederá a generar las autocorrelaciones de la muestra, dichos resultados se observan en la figura 15. En dicha figura se observa que la autocorrelación en el desfase 6 llega a los límites de significancia, por lo que se puede considerar como significativa.

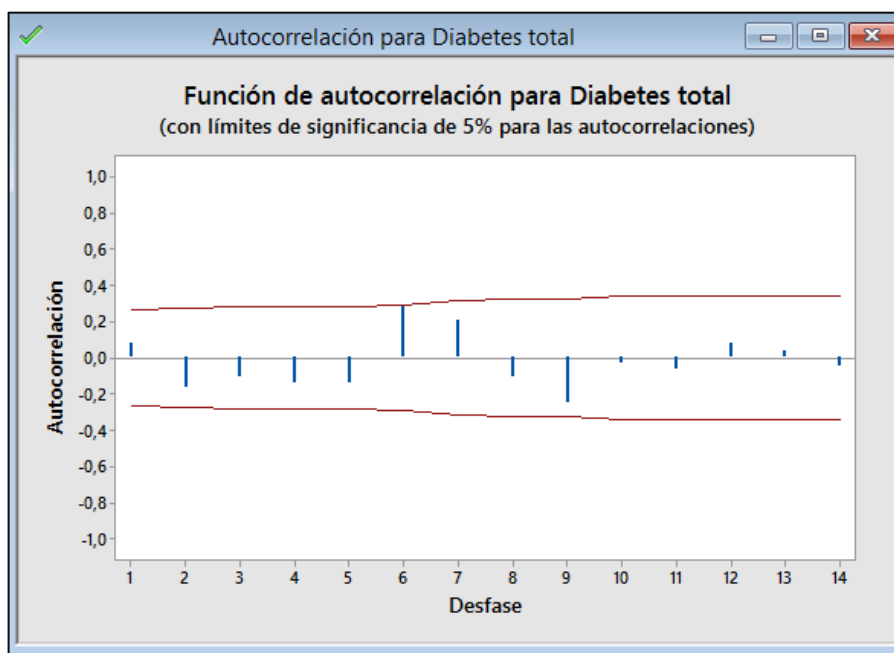


Figura 15. Autocorrelaciones para los datos de la diabetes total

Tomando en cuenta el desfase 6 como autocorrelación significativa de la muestra, tenemos una autocorrelación de 0.29, tal como se observa en la figura 16, por lo que a continuación se procede a verificar el nivel de significancia.

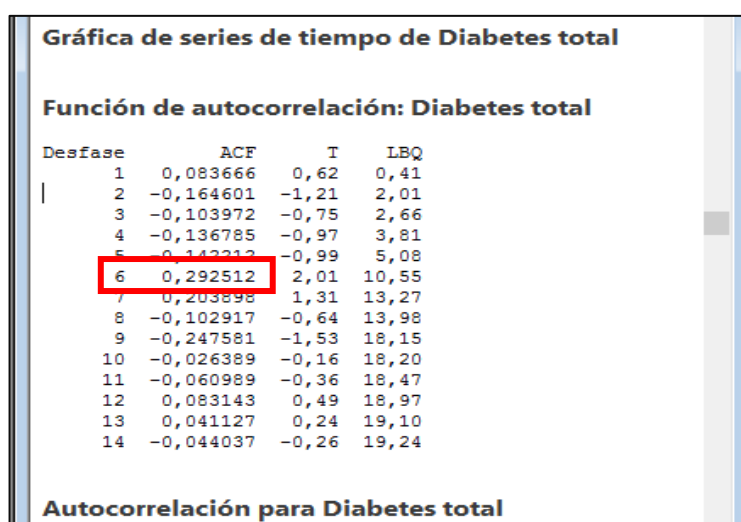


Figura 16. Desfases de la función de autocorrelación.

3.4.3 Pruebas de hipótesis referidas a las distribuciones de frecuencias

Prueba Q (chi cuadrada).- Para este punto se especifican las siguientes hipótesis:

- ✓ H_0 : El grupo de las autocorrelaciones se considera significativamente nulo (ceranos a cero).

✓ H_1 : El grupo de las autocorrelaciones no es significativamente nulo.

El valor de Q (LBQ) para 14 retrasos de tiempo es de 19,24 (de Minitab), como se indica en la figura 16. El probabilidad 0.05 de una distribución chi-cuadrada con 14 grados de libertad es 23,68 (véase la tabla 4). Puesto que 19,24 es menor que 23,68, no se puede rechazar H_0 con grado de significancia de 5%.

v/p	0,001	0,0025	0,005	0,01	0,025	0,05	0,1	0,15	0,2
1	10,8274	9,1404	7,8794	6,6349	5,0239	3,8415	2,7055	2,0722	1,6424
2	13,8150	11,9827	10,5965	9,2104	7,3778	5,9915	4,6052	3,7942	3,2189
3	16,2660	14,3202	12,8381	11,3449	9,3484	7,8147	6,2514	5,3170	4,6416
4	18,4662	16,4238	14,8602	13,2767	11,1433	9,4877	7,7794	6,7449	5,9886
5	20,5147	18,3854	16,7496	15,0863	12,8325	11,0705	9,2363	8,1152	7,2893
6	22,4575	20,2491	18,5475	16,8119	14,4494	12,5916	10,6446	9,4461	8,5581
7	24,3213	22,0402	20,2777	18,4753	16,0128	14,0671	12,0170	10,7479	9,8032
8	26,1239	23,7742	21,9549	20,0902	17,5345	15,5073	13,3616	12,0271	11,0301
9	27,8767	25,4625	23,5893	21,6660	19,0228	16,9190	14,6837	13,2880	12,2421
10	29,5879	27,1119	25,1881	23,2093	20,4832	18,3070	15,9872	14,5339	13,4420
11	31,2635	28,7291	26,7569	24,7250	21,9200	19,6752	17,2750	15,7671	14,6314
12	32,9092	30,3182	28,2997	26,2170	23,3367	21,0261	18,5493	16,9893	15,8120
13	34,5274	31,8830	29,8193	27,6882	24,7356	22,3620	19,8119	18,2020	16,9848
14	36,1239	33,4262	31,3194	29,1412	26,1189	23,6848	21,0641	19,4062	18,1508
15	37,6978	34,9494	32,8015	30,5780	27,4884	24,9958	22,3071	20,6030	19,3107
16	39,2518	36,4555	34,2671	31,9999	28,8453	26,2962	23,5418	21,7931	20,4651
17	40,7911	37,9462	35,7184	33,4087	30,1910	27,5871	24,7690	22,9770	21,6146

Tabla 4 Distribución Chi cuadrado, en donde p= probabilidad de encontrar un valor mayor o igual que el chi cuadrado tabulado y v= grados de libertad. (Condor, 2009)

Prueba t: Para este punto se especifican las siguientes hipótesis:

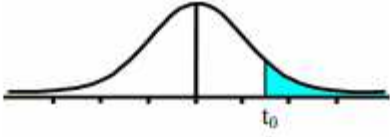
✓ H_0 : El valor de las autocorrelaciones se considera significativamente nulo.

✓ H_1 : El valor de las autocorrelaciones no es significativamente nulo.

El valor del estadístico de prueba t es equivalente a 2,01 (de Minitab) tal como lo muestra la figura 16, para los valores críticos del estadístico t con $df = n - 1$, de la tabla t (véase la tabla 5). Se puede observar que $\pm 2,0$ son los puntos superior e inferior, de una distribución t con 13 grados de libertad (14 datos), para un nivel de significancia de 5%

(0.025). Esto nos indica que no se puede aceptar la hipótesis nula con el nivel de significancia de 5%, ya que $2.0 < 2.01$, es decir, los datos no son aleatorios.

Tabla t-Student



Grados de libertad	0.25	0.1	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453

Tabla 5 Distribución t (Condor, 2009)

Haciendo una comparación entre estas dos pruebas, la prueba de distribución t muestra que los datos obtenidos no son aleatorios y se puede comenzar a construir el modelo.

3.4.4 Construcción y evaluación del modelo

Modelos ARIMA

En la figura 17 se despliega el ajuste del modelo ARIMA (0,0,0) (0,1,1) en la herramienta Minitab.

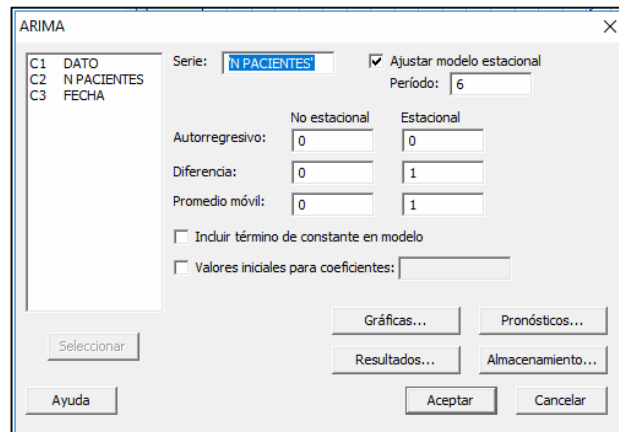


Figura 17. Ventana de diálogo del Minitab para los ajustes del modelo ARIMA (0,0 0) (0,1,1)

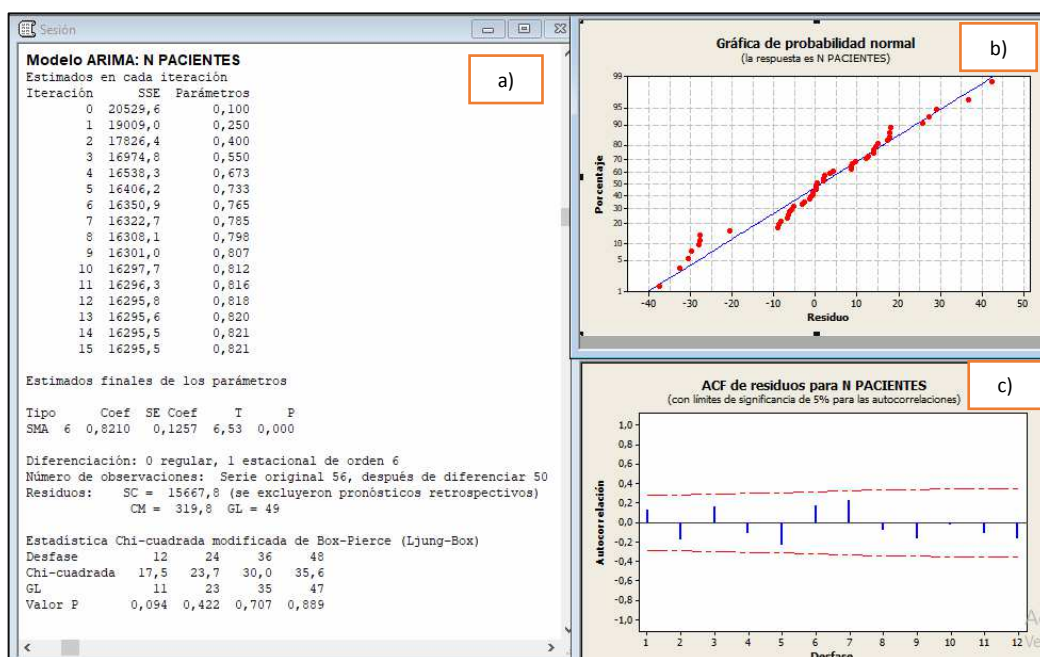


Figura 18. Gráficas de Probabilidad normal y auto correlación de residuos ARIMA (0,0,0) (0,1,1)

En los resultados del modelo de la figura 18(a) se observa que el modelo ARIMA (0,0,0) (0,1,1) con diferenciación en 1 estacional en el desfase 6 nos proyecta un error cuadrado medio de $MC = 319,8$ con 49 grados de libertad, y la hipótesis nula indica que los datos siguen una distribución normal. Debido a que el valor p de 0.889 es mayor al nivel de significancia de 0.05, la decisión es no rechazar la hipótesis nula; esto se puede confirmar con el análisis de la gráfica de probabilidad normal, tal como se muestra en el gráfico 18(b), en el cual los puntos siguen la línea ajustada adecuadamente, esto indica que la distribución normal parece ajustarse apropiadamente a los datos. En el gráfico 18(c) de autocorrelación de

residuos se puede indicar que se encuentran dentro de los límites de significancia por tal motivo es aceptable el modelo.

La figura 19 presenta el ajuste del modelo ARIMA (0,0,0) (1,1,0) en la herramienta Minitab, con periodos de estacionalidad de 6 desfases.

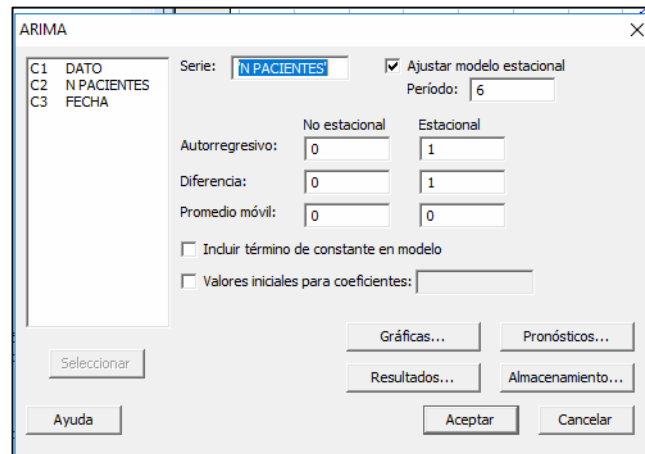


Figura 19. Ventana de diálogo del Minitab para los ajustes del modelo ARIMA (0,0,0) (1,1,0)

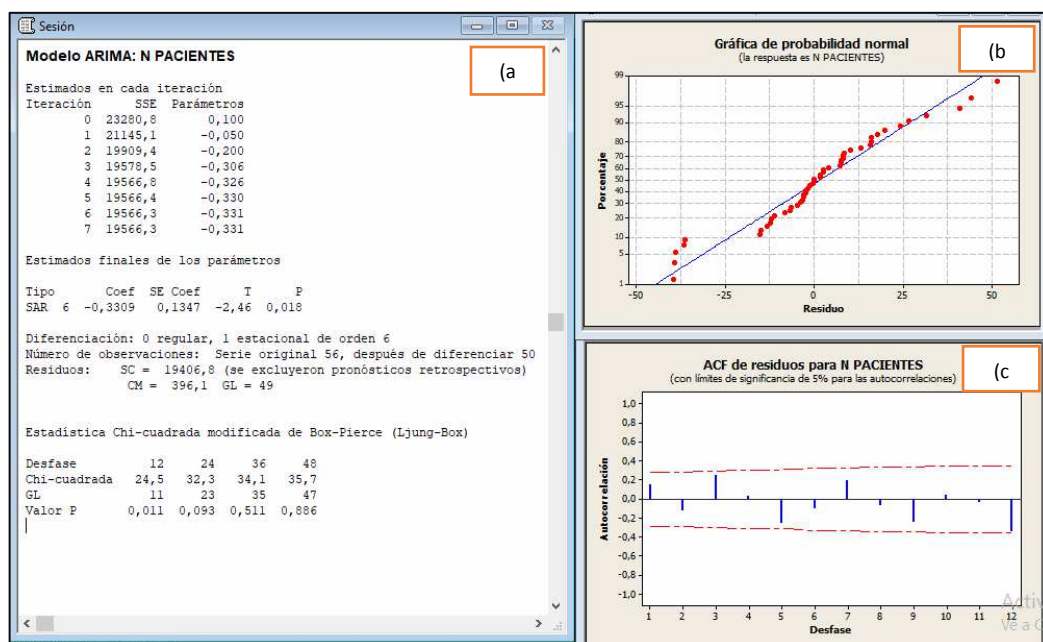


Figura 20. Gráficas de Probabilidad normal y auto correlación de residuos ARIMA (0,0,0) (1,1,0)

En los resultados del modelo de la figura 20(a se observa que el modelo ARIMA (0,0,0) (1,1,0) con diferenciación en 1 estacional en desfases de 6 nos proyecta un error cuadrado medio de MC= 396,1 con 49 grados de libertad, y la hipótesis nula indica que los datos siguen una distribución normal.

Debido a que el valor p de 0.886 es mayor al nivel de significancia de 0.05, la decisión es no rechazar la hipótesis nula; esto se puede confirmar con el análisis de la gráfica de probabilidad normal, tal como se muestra en el gráfico 20(b), en el cual los puntos siguen la línea ajustada adecuadamente, de esta manera la distribución normal parece ajustarse apropiadamente a los datos.

En el gráfico 20(c) de autocorrelación de residuos se indica que se encuentran dentro de los límites de significancia por lo tanto este modelo también es aceptable.

La figura 21 presenta el ajuste del modelo ARIMA (0,0,1) (0,1,0) en la herramienta Minitab, con periodos de estacionalidad de 6 desfases.

Figura 21. Ventana de diálogo del Minitab para los ajustes del modelo ARIMA (0,0,1) (0,1,0)

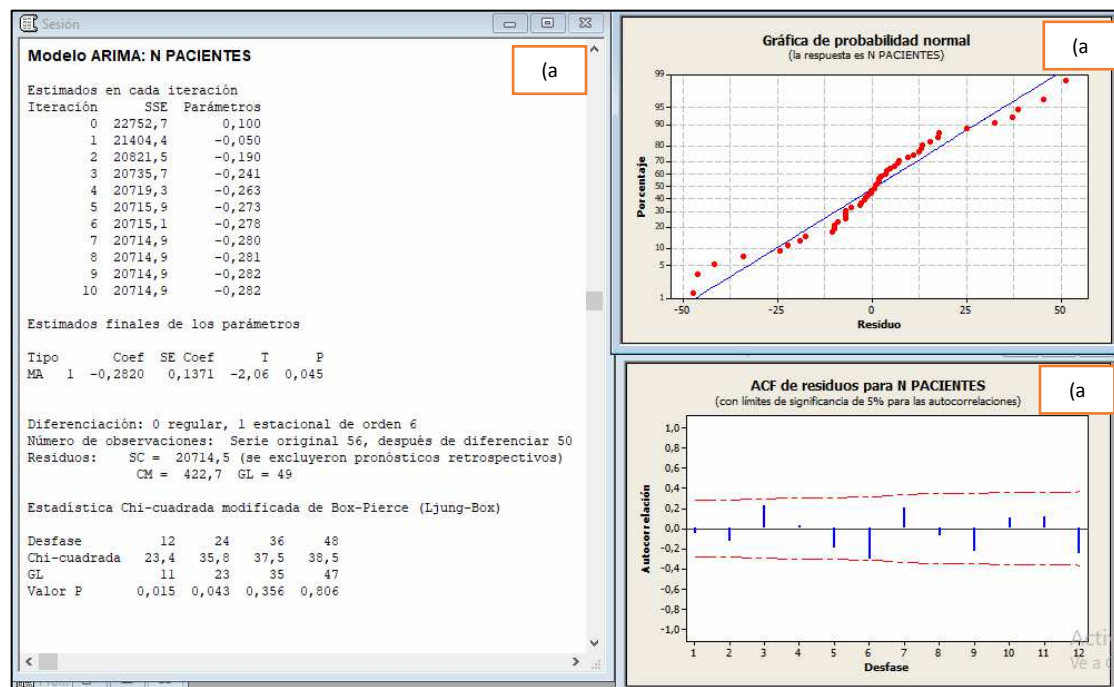


Figura 22. Gráficas de Probabilidad normal y auto correlación de residuos ARIMA (0,0,1) (0,1,0)

En los resultados del modelo de la figura 22(a se observa que el modelo ARIMA (0,0,1) (0,1,0) con diferenciación en 1 estacional en desfases de 6, nos proyecta un error cuadrado medio de MC= 422,7 con 49 grados de libertad, y la hipótesis nula indica que los datos siguen una distribución normal. Debido a que el valor p de 0.806 es mayor al nivel de significancia de 0.05, la decisión es no rechazar la hipótesis nula; esto se puede confirmar con el análisis de la gráfica de probabilidad normal, tal como se muestra en el gráfico 22(b, en el cual los puntos siguen la línea ajustada adecuadamente, de esta manera la distribución normal parece ajustarse apropiadamente a los datos. El gráfico 22(c de autocorrelación de residuos indica que se encuentran dentro de los límites de significancia por lo tanto este modelo también es aceptable.

3.4.5 Resultados

Dado que los tres modelos de ARIMA analizados se ajustan muy bien a los datos, se procede al análisis de los errores cuadráticos Medios de los modelos planteados, cuyos valores son:

ARIMA (0,0,0) (0,1,1) => MC = 319.8

ARIMA (0,0,0) (1,1,0) => MC = 396.1

ARIMA (0,0,1) (0,1,0) => MC = 422.7

El modelo que más se ajusta a los datos analizados es el modelo ARIMA (0,0,0) (0,1,1) debido a que proyecta la menor cantidad de errores posibles al momento de realizar los pronósticos.

Este modelo se puede tomar como referencia para realizar pronósticos de diferentes enfermedades que aquejan a la ciudad de Manta en el Instituto Ecuatoriano de Seguridad Social y en otras Instituciones como Hospitales y clínicas, debido a que tener un número probable de pacientes permitiría tener un stock aceptable de medicamentos y disponibilidad de personal médico que atiendan a los pacientes con prontitud y así mejorar el sistema de salud en dicha ciudad.

3.4.6 Aplicativo Web en R

En la figura 23 podemos observar la pantalla principal del aplicativo web desarrollado en R, el cual permite modelar los distintos modelos ARIMA y poder analizarlos en tiempo real.

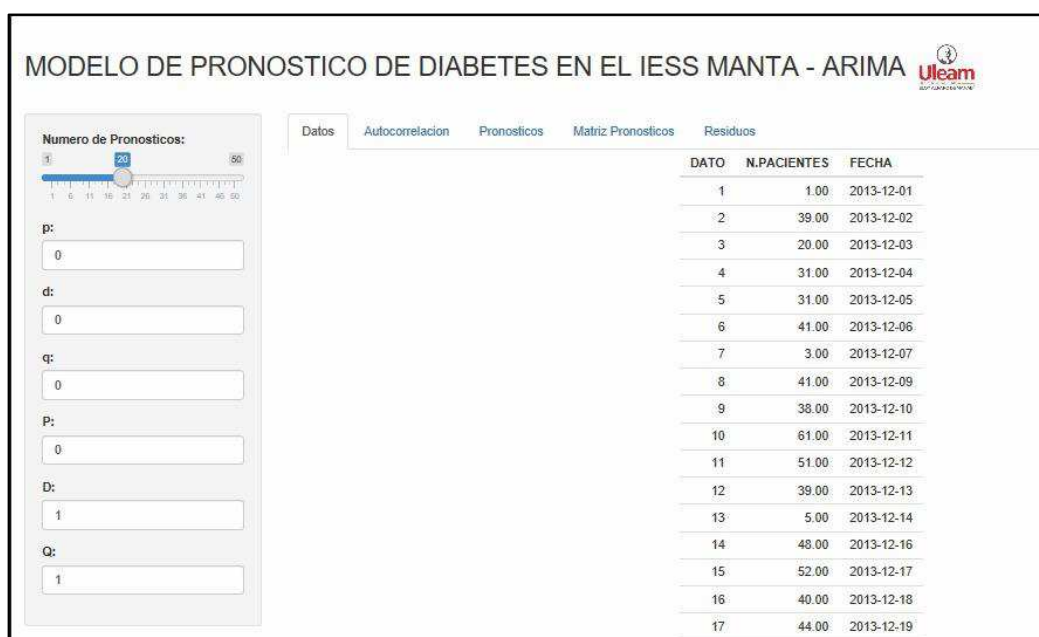


Figura 23. Pantalla Principal del aplicativo Web de Modelo de Pronóstico ARIMA

En la figura 24 se puede observar la pantalla de autocorrelación, la cual muestra el gráfico de autocorrelación de los datos utilizados en donde podemos realizar el análisis de la serie de tiempo que muestran los datos y determinar la estacionalidad de los datos.

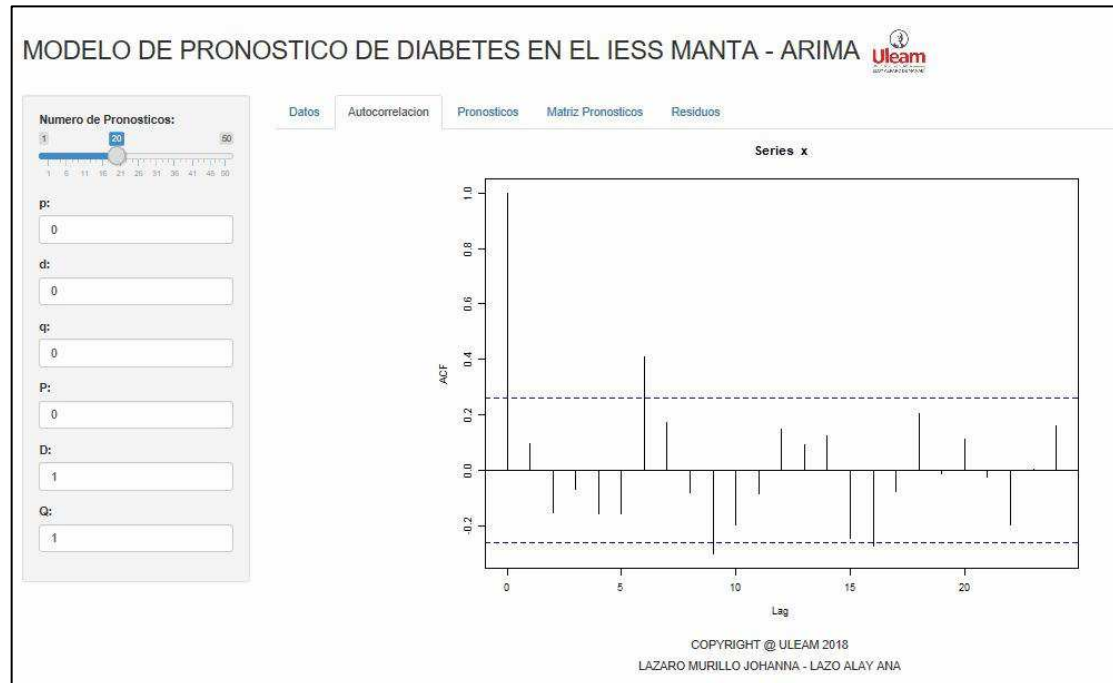


Figura 24 Pantalla de Autocorrelación de los datos del aplicativo Web de Modelo de Pronóstico ARIMA

En la figura 25 se observar la pantalla de pronóstico, la cual muestra el gráfico de serie de tiempo de los datos con los pronósticos generados.

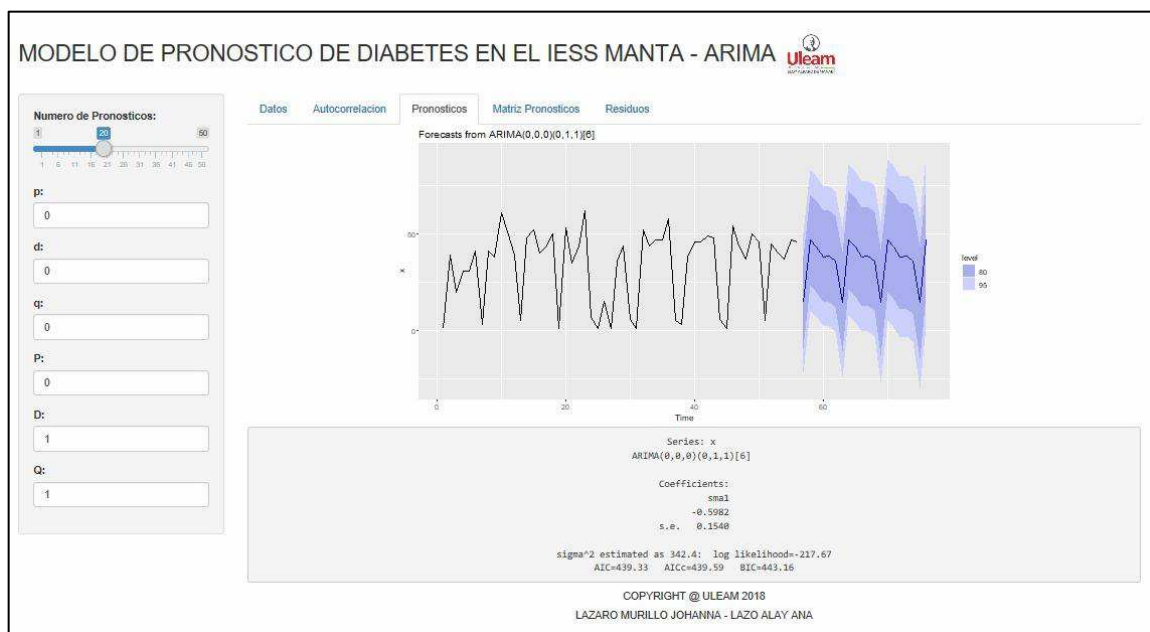


Figura 25. Pantalla de la gráfica de Pronósticos de los datos del aplicativo Web de Modelo de Pronóstico ARIMA

En la figura 26 podemos observar la pantalla de la matriz de pronósticos, la cual muestra los pronósticos generados por el modelo seleccionado.



Figura 26. Pantalla de la Matriz de Pronósticos de los datos del aplicativo Web de Modelo de Pronóstico ARIMA

En la figura 27 se puede observar la pantalla de residuos, la cual muestra los residuos de los pronósticos generados por el modelo seleccionado, en esta pantalla tenemos cuatro gráficos, el gráfico de distribución normal, la serie de tiempo de los residuos generados, la gráfica de autocorrelación de los residuos y la campana de la distribución normal; que nos ayudan a determinar si el modelo seleccionado es aceptable.

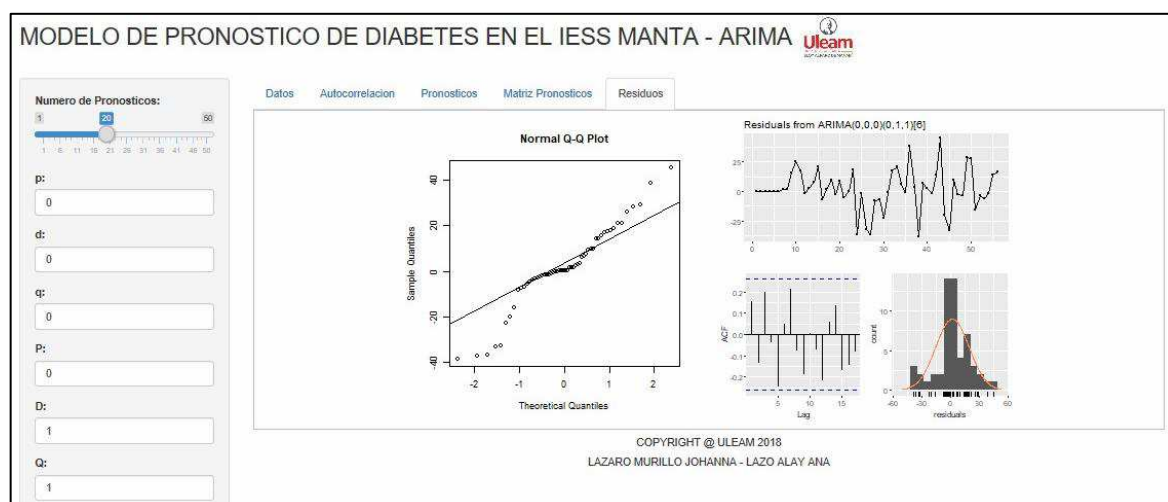
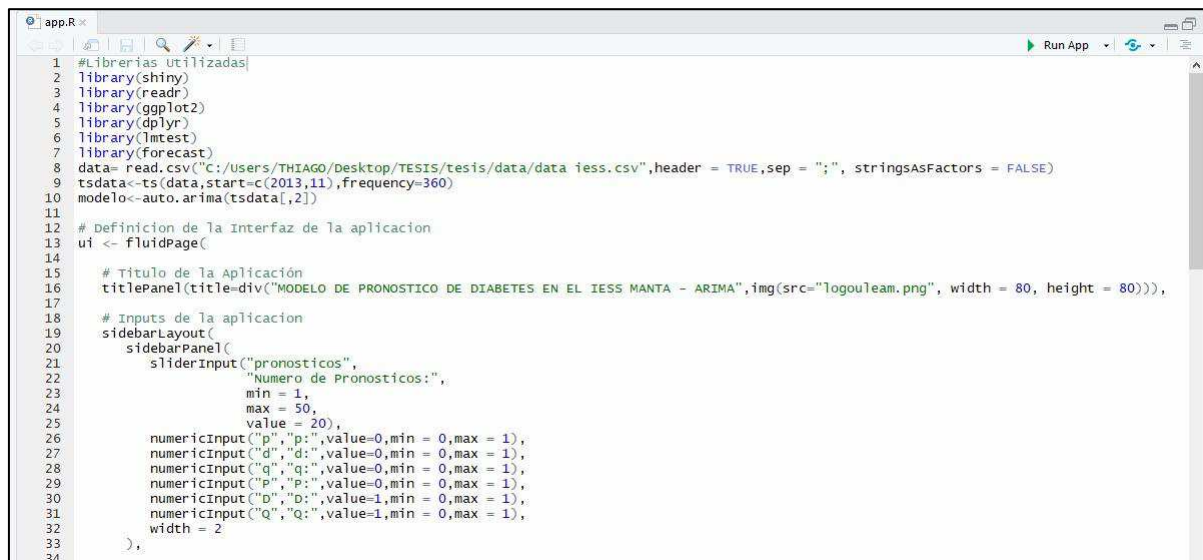


Figura 27. Pantalla de la gráfica de Residuos de los Modelos de Pronóstico ARIMA

La aplicación web en R está conformada por dos partes: la interfaz gráfica o ui y la parte de servidor o server, en la figura 28 podemos observar la declaración de las librerías utilizadas para el desarrollo del aplicativo web, además de la creación de la interfaz gráfica de la aplicación; es decir la ui en donde se declaran las variables de ingreso para los parámetros del modelo ARIMA (p,d,q) (P,D,Q).

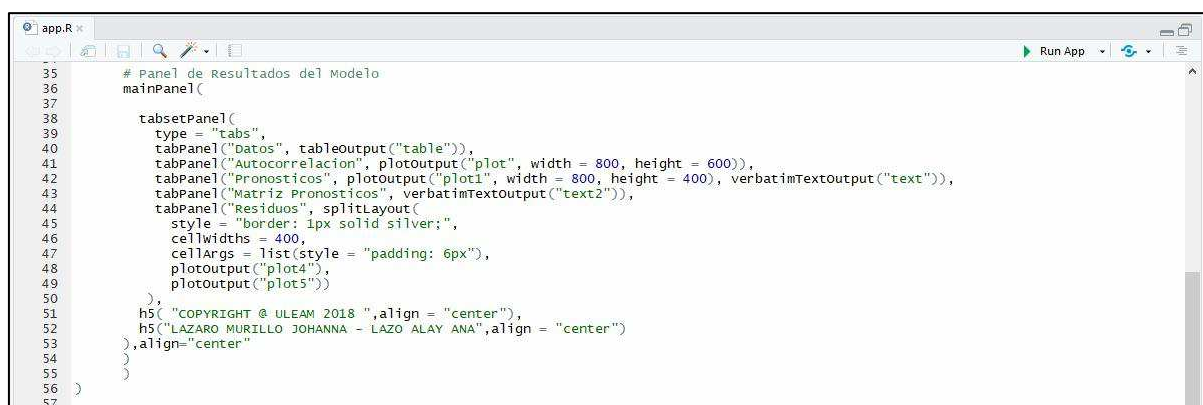


```

1 #Librerías utilizadas
2 library(shiny)
3 library(readr)
4 library(ggplot2)
5 library(dplyr)
6 library(lmtest)
7 library(Forecast)
8 data= read.csv("C:/Users/THIAGO/Desktop/TESIS/tesis/data/data_ies.csv",header = TRUE,sep = ";", stringsAsFactors = FALSE)
9 tsdata<-ts(data,start=c(2013,11),frequency=360)
10 modelo<-auto.arima(tsdata[,2])
11
12 # Definición de la Interfaz de la aplicación
13 ui <- fluidPage(
14   # Título de la Aplicación
15   titlePanel(title=div("MODELO DE PRONOSTICO DE DIABETES EN EL IEES MANTA - ARIMA",img(src="logouleam.png", width = 80, height = 80))),
16   # Inputs de la aplicación
17   sidebarLayout(
18     sidebarPanel(
19       sliderInput("pronosticos",
20         "Numero de Pronosticos:",
21         min = 1,
22         max = 50,
23         value = 20),
24       numericInput("p","p:",value=0,min = 0,max = 1),
25       numericInput("d","d:",value=0,min = 0,max = 1),
26       numericInput("q","q:",value=0,min = 0,max = 1),
27       numericInput("P","P:",value=0,min = 0,max = 1),
28       numericInput("D","D:",value=1,min = 0,max = 1),
29       numericInput("Q","Q:",value=1,min = 0,max = 1),
30       width = 2
31     ),
32   )
33 )
34
  
```

Figura 28. Declaración de librerías y variables input del aplicativo Web de Modelo de Pronóstico ARIMA.

En la figura 29 podemos observar la declaración de las variables de salida en la interfaz gráfica (ui), que es donde se organiza la presentación de los resultados del modelo ARIMA.



```

35 # Panel de Resultados del Modelo
36 mainPanel(
37   tabsetPanel(
38     type = "tabs",
39     tabPanel("Datos", tableoutput("table")),
40     tabPanel("Autocorrelacion", plotoutput("plot", width = 800, height = 600)),
41     tabPanel("Pronosticos", plotoutput("plot1", width = 800, height = 400), verbatimTextoutput("text")),
42     tabPanel("Matriz Pronosticos", verbatimTextoutput("text2")),
43     tabPanel("Residuos", splitLayout(
44       style = "border: 1px solid silver;",
45       cellwidths = 400,
46       cellargs = list(style = "padding: 6px"),
47       plotoutput("plot4"),
48       plotoutput("plot5")
49     )
50   ),
51   h5("COPYRIGHT @ ULEAM 2018 ",align = "center"),
52   h5("LAZARO MURILLO JOHANNA - LAZO ALAY ANA",align = "center")
53 ),align="center"
54 )
55 )
56 )
57
  
```

Figura 29. Declaración de las variables output del aplicativo Web de Modelo de Pronóstico ARIMA.

En la figura 30 podemos observar la parte del servidor de la aplicación web, en la cual se recogen los datos de ingreso, luego se procesan y se obtienen los datos que se van a mostrar en el ui de la aplicación.

```

58 # Definición del servidor
59 server <- function(input, output) {
60   data[,2]<-as.numeric(data[, 2])
61   x<-data[,2]
62
63   output$plot <- renderPlot({
64     x<-ts(x)
65     acf(x,24)
66   })
67   output$plot1 <- renderPlot({
68     x<-ts(x)
69     fit<-Arima(x,order=c(input$p,input$d,input$q),seasonal = list(order = c(input$p, input$d, input$q), period = 6))
70     fit %>% forecast(h=input$pronosticos) %>% autoplot(include=80)
71   })
72   output$text <- renderPrint({
73     x<-ts(x)
74     fit<-Arima(x,order=c(input$p,input$d,input$q),seasonal = list(order = c(input$p, input$d, input$q), period = 6))
75     fit
76   })
77   output$text2 <- renderPrint({
78     x<-ts(x)
79     fit<-Arima(x,order=c(input$p,input$d,input$q),seasonal = list(order = c(input$p, input$d, input$q), period = 6))
80     forecast(fit,h=input$pronosticos)
81   })
82   output$plot4 <- renderPlot({
83     x<-ts(x)
84     fit<-Arima(x,order=c(input$p,input$d,input$q),seasonal = list(order = c(input$p, input$d, input$q), period = 6))
85     qqnorm(residuals(fit))
86     qqline(residuals(fit))
87   })
88   output$plot5 <- renderPlot({
89     x<-ts(x)
90     fit<-Arima(x,order=c(input$p,input$d,input$q),seasonal = list(order = c(input$p, input$d, input$q), period = 6))
91     checkresiduals(fit)
92   })
93   output$table <- renderTable({ data })
94 }
95
  
```

Figura 30. Declaración de las variables output del aplicativo Web de Modelo de Pronóstico ARIMA

En la figura 31 podemos observar la declaración en la librería shiny tanto la interfaz gráfica el ui y el servidor, de esta manera se ejecuta la aplicación.

```

96
97 # Run de la aplicación
98 shinyApp(ui = ui, server = server)
99
  
```

Ilustración 31. Declaración de las variables output del aplicativo Web de Modelo de Pronóstico ARIMA

CAPITULO

IV

CONCLUSIONES Y RECOMENDACIONES

CAPÍTULO IV

CONCLUSIONES

Se utilizó la metodología ARIMA para desarrollar un modelo de pronóstico para determinar el número futuro de pacientes con Diabetes que atendería el IESS del Cantón Manta, dando una cantidad promedio de 33 pacientes con Diabetes que se atenderían por día.

En la preparación de los datos para aplicar los modelos de pronóstico de serie de tiempo ARIMA se obtuvo que los datos recopilados reflejan una pequeña estacionalidad en periodos de seis retrasos cuyo resultado permitió determinar los valores apropiados de los parámetros (p,d,q) y (P,D,Q) para el modelo ARIMA, de esta manera los modelos candidatos fueron: ARIMA(0,0,0; 0,1,1) - ARIMA(0,0,0; 1,1,0) - ARIMA(0,0,1; 0,1,0) y, para determinar si estos modelos se ajustaban a los datos, se procedió a aplicar técnicas estadísticas tales como las pruebas de hipótesis mediante la prueba Q (chi-cuadrada) y la prueba t.

De esta manera los tres modelos candidatos se obtuvieron residuos no correlacionados, y además los errores cuadrados medios determinaron que el modelo más óptimo es el modelo ARIMA (0,0,0; 0,1,1) con menor error cuadrático medio.

Se procedió a realizar un aplicativo web en R para poder observar los resultados de los residuos de manera gráfica y poder analizarlos, para demostrar el comportamiento de los datos con los modelos de serie de tiempo ARIMA candidatos en tiempo real, de esta manera se concluye que R es un lenguaje de programación muy flexible y efectivo a la hora de desarrollar modelos de pronósticos puesto que es sencillo de utilizar además de ser una herramienta estadística muy precisa.

RECOMENDACIONES

- ✓ Incentivar a las diferentes instituciones a la utilización de métodos de pronósticos para ayudar a mejorar el servicio que se ofrece al público.
- ✓ Se recomienda a las instituciones proporcionar los datos para su respectivo análisis estadístico y modelado de los mismos.
- ✓ El trabajo realizado se puede complementar con la obtención de los datos más recientes, para actualizar el modelo resultante de esta investigación.
- ✓ Como última recomendación se sugiere repetir el ejercicio con datos de otros tipos de enfermedades usuales en la ciudad de Manta y si es posible extenderse a otras ciudades.

Referencias bibliográficas

- A.D.A.M. (14 de Agosto de 2018). *MedlinPlus*. Obtenido de <https://medlineplus.gov/spanish/ency/article/001214.htm>
- Cabena, P., Verhees, J., Zanasi, A., & Hadjnia, P. (1998). *Discovering Data Mining: From Concept to Implementation*. Upper Saddle River, Nueva Jersey, Estado Unidos: Prentice-Hall.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*.
- Condor, I. (Enero de 2009). *Aula Clic*. Obtenido de http://www.aulaclic.es/minitab/t_1_1.htm
- Data Carpentry. (2018). Obtenido de <https://datacarpentry.org/R-ecology-lesson/00-before-we-start.html>
- Fernández, S. d. (1 de Febrero de 2016). *Modelo ARIMA (p,d,q) (P,D,Q)s*. Obtenido de <http://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf>
- García, E. (03 de Septiembre de 2014). *Slideshare.net*. Obtenido de https://es.slideshare.net/ector_03/ap2-51204100
- González, V. (2014). Aspectos críticos del empleo en salud de modelos estadísticos de clasificación. *Rev Méd Electrón [revista en Internet]*, 36.
- Hanke, J. E., & Wichern, D. W. (2010). *Pronósticos en los negocios* (9a. Edición ed.). México: Pearson Prentice Hall.
- Hu, F. B., Manson, J. E., Stampfer, M. J., Colditz, G., & Willett, W. C. (2001). Diet, Lifestyle, and the Risk of Type 2 Diabetes Mellitus in Women. *New England Journal of Medicine*, 345: 790 - 797.
- IBM. (Agosto de 2017). *IBM Knowledge Center*. Obtenido de IBM Knowledge Center.
- Ihaka, R., & Gentleman, R. (1996). *R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical* (Vol. 5).
- INEC. (s.f.). *Instituto Nacional de Estadísticas y Censos*. Recuperado el 09 de Enero de 2018, de http://www.ecuadorencifras.gob.ec/documentos/web-inec/Poblacion_y_Demografia/Nacimientos_Defunciones/2016/
- Larose, D. T. (2005). *Discovering Knowledge in Data, an introduction to data mining*. John Wiley & Sons.
- Mendivil, C., & Sierra, I. (2011). *Hacia el manejo practico de la Diabetes Mellitus* (Vol. 3). Colombia: Burgos.
- Microsoft Corporation. (Agosto de 2013). *MSDN Library*. Obtenido de <http://msdn.microsoft.com/eses/library/cc645779.aspx#feedback>
- Weiss, S., & Indurkha, N. (1998). *Predictive Data Mining. A Practical Guide*. San Francisco: Morgan Kaufmann.

ANEXOS

CERTIFICACIÓN

Por medio de la presente yo, Ing. Oscar Armando González López, técnico informático de la Subdirección de Salud del Instituto Ecuatoriano de Seguridad Social (IESS) de Manabí, durante el periodo de septiembre del 2011 a Julio del 2014; certifico que los datos utilizados en este trabajo de titulación son reales, tomados de la base de datos del IESS de Manabí, la cual se ha facilitado en archivo .csv el mismo que contiene datos de Octubre del 2013 a Enero del 2014.

Ing. Oscar Armando González López

Manta, 06 de Septiembre del 2018

TABLA 3-Distribución Chi Cuadrado χ^2

P = Probabilidad de encontrar un valor mayor o igual que el chi cuadrado tabulado, v = Grados de Libertad

v/p	0,001	0,0025	0,005	0,01	0,025	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
1	10,8274	9,1404	7,8794	6,6349	5,0239	3,8415	2,7055	2,0722	1,6424	1,3233	1,0742	0,8735	0,7083	0,5707	0,4549
2	13,8150	11,9827	10,5965	9,2104	7,3778	5,9915	4,6052	3,7942	3,2189	2,7726	2,4079	2,0996	1,8326	1,5970	1,3863
3	16,2660	14,3202	12,8381	11,3449	9,3484	7,8147	6,2514	5,3170	4,6416	4,1083	3,6649	3,2831	2,9462	2,6430	2,3660
4	18,4662	16,4238	14,8602	13,2767	11,1433	9,4877	7,7794	6,7449	5,9886	5,3853	4,8784	4,4377	4,0446	3,6871	3,3567
5	20,5147	18,3854	16,7496	15,0863	12,8325	11,0705	9,2363	8,1152	7,2893	6,6257	6,0644	5,5731	5,1319	4,7278	4,3515
6	22,4575	20,2491	18,5475	16,8119	14,4494	12,5916	10,6446	9,4461	8,5581	7,8408	7,2311	6,6948	6,2108	5,7652	5,3481
7	24,3213	22,0402	20,2777	18,4753	16,0128	14,0671	12,0170	10,7479	9,8032	9,0371	8,3834	7,8061	7,2832	6,8000	6,3458
8	26,1239	23,7742	21,9549	20,0902	17,5345	15,5073	13,3616	12,0271	11,0301	10,2189	9,5245	8,9094	8,3505	7,8325	7,3441
9	27,8767	25,4625	23,5893	21,6660	19,0228	16,9190	14,6837	13,2880	12,2421	11,3887	10,6564	10,0060	9,4136	8,8632	8,3428
10	29,5879	27,1119	25,1881	23,2093	20,4832	18,3070	15,9872	14,5339	13,4420	12,5489	11,7807	11,0971	10,4732	9,8922	9,3418
11	31,2635	28,7291	26,7569	24,7250	21,9200	19,6752	17,2750	15,7671	14,6314	13,7007	12,8987	12,1836	11,5298	10,9199	10,3410
12	32,9092	30,3182	28,2997	26,2170	23,3367	21,0261	18,5493	16,9893	15,8120	14,8454	14,0111	13,2661	12,5838	11,9463	11,3403
13	34,5274	31,8830	29,8193	27,6882	24,7356	22,3620	19,8119	18,2020	16,9848	15,9839	15,1187	14,3451	13,6356	12,9717	12,3398
14	36,1239	33,4262	31,3194	29,1412	26,1189	23,6848	21,0641	19,4062	18,1508	17,1169	16,2221	15,4209	14,6853	13,9961	13,3393
15	37,6978	34,9494	32,8015	30,5780	27,4884	24,9958	22,3071	20,6030	19,3107	18,2451	17,3217	16,4940	15,7332	15,0197	14,3389
16	39,2518	36,4555	34,2671	31,9999	28,8453	26,2962	23,5418	21,7931	20,4651	19,3689	18,4179	17,5646	16,7795	16,0425	15,3385
17	40,7911	37,9462	35,7184	33,4087	30,1910	27,5871	24,7690	22,9770	21,6146	20,4887	19,5110	18,6330	17,8244	17,0646	16,3382
18	42,3119	39,4220	37,1564	34,8052	31,5264	28,8693	25,9894	24,1555	22,7595	21,6049	20,6014	19,6993	18,8679	18,0860	17,3379
19	43,8194	40,8847	38,5821	36,1908	32,8523	30,1435	27,2036	25,3289	23,9004	22,7178	21,6891	20,7638	19,9102	19,1069	18,3376
20	45,3142	42,3358	39,9969	37,5663	34,1696	31,4104	28,4120	26,4976	25,0375	23,8277	22,7745	21,8265	20,9514	20,1272	19,3374
21	46,7963	43,7749	41,4009	38,9322	35,4789	32,6706	29,6151	27,6620	26,1711	24,9348	23,8578	22,8876	21,9915	21,1470	20,3372
22	48,2676	45,2041	42,7957	40,2894	36,7807	33,9245	30,8133	28,8224	27,3015	26,0393	24,9390	23,9473	23,0307	22,1663	21,3370
23	49,7276	46,6231	44,1814	41,6383	38,0756	35,1725	32,0069	29,9792	28,4288	27,1413	26,0184	25,0055	24,0689	23,1852	22,3369
24	51,1790	48,0336	45,5584	42,9798	39,3641	36,4150	33,1962	31,1325	29,5533	28,2412	27,0960	26,0625	25,1064	24,2037	23,3367
25	52,6187	49,4351	46,9280	44,3140	40,6465	37,6525	34,3816	32,2825	30,6752	29,3388	28,1719	27,1183	26,1430	25,2218	24,3366
26	54,0511	50,8291	48,2898	45,6416	41,9231	38,8851	35,5632	33,4295	31,7946	30,4346	29,2463	28,1730	27,1789	26,2395	25,3365
27	55,4751	52,2152	49,6450	46,9628	43,1945	40,1133	36,7412	34,5736	32,9117	31,5284	30,3193	29,2266	28,2141	27,2569	26,3363
28	56,8918	53,5939	50,9936	48,2782	44,4608	41,3372	37,9159	35,7150	34,0266	32,6205	31,3909	30,2791	29,2486	28,2740	27,3362
29	58,3006	54,9662	52,3355	49,5878	45,7223	42,5569	39,0875	36,8538	35,1394	33,7109	32,4612	31,3308	30,2825	29,2908	28,3361

TABLA 3-Distribución Chi Cuadrado χ^2 . (Continuación)

v/p	0,001	0,0025	0,005	0,01	0,025	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
30	59,7022	56,3325	53,6719	50,8922	46,9792	43,7730	40,2560	37,9902	36,2502	34,7997	33,5302	32,3815	31,3159	30,3073	29,3360
31	61,0980	57,6921	55,0025	52,1914	48,2319	44,9853	41,4217	39,1244	37,3591	35,8871	34,5981	33,4314	32,3486	31,3235	30,3359
32	62,4873	59,0461	56,3280	53,4857	49,4804	46,1942	42,5847	40,2563	38,4663	36,9730	35,6649	34,4804	33,3809	32,3394	31,3359
33	63,8694	60,3953	57,6483	54,7754	50,7251	47,3999	43,7452	41,3861	39,5718	38,0575	36,7307	35,5287	34,4126	33,3551	32,3358
34	65,2471	61,7382	58,9637	56,0609	51,9660	48,6024	44,9032	42,5140	40,6756	39,1408	37,7954	36,5763	35,4438	34,3706	33,3357
35	66,6192	63,0760	60,2746	57,3420	53,2033	49,8018	46,0588	43,6399	41,7780	40,2228	38,8591	37,6231	36,4746	35,3858	34,3356
36	67,9850	64,4097	61,5811	58,6192	54,4373	50,9985	47,2122	44,7641	42,8788	41,3036	39,9220	38,6693	37,5049	36,4008	35,3356
37	69,3476	65,7384	62,8832	59,8926	55,6680	52,1923	48,3634	45,8864	43,9782	42,3833	40,9839	39,7148	38,5348	37,4156	36,3355
38	70,7039	67,0628	64,1812	61,1620	56,8955	53,3835	49,5126	47,0072	45,0763	43,4619	42,0450	40,7597	39,5643	38,4302	37,3354
39	72,0550	68,3830	65,4753	62,4281	58,1201	54,5722	50,6598	48,1263	46,1730	44,5395	43,1053	41,8040	40,5935	39,4446	38,3354
40	73,4029	69,6987	66,7660	63,6908	59,3417	55,7585	51,8050	49,2438	47,2685	45,6160	44,1649	42,8477	41,6222	40,4589	39,3353
45	80,0776	76,2229	73,1660	69,9569	65,4101	61,6562	57,5053	54,8105	52,7288	50,9849	49,4517	48,0584	46,7607	45,5274	44,3351
50	86,6603	82,6637	79,4898	76,1538	71,4202	67,5048	63,1671	60,3460	58,1638	56,3336	54,7228	53,2576	51,8916	50,5923	49,3349
55	93,1671	89,0344	85,7491	82,2920	77,3804	73,3115	68,7962	65,8550	63,5772	61,6650	59,9804	58,4469	57,0160	55,6539	54,3348
60	99,6078	95,3443	91,9518	88,3794	83,2977	79,0820	74,3970	71,3411	68,9721	66,9815	65,2265	63,6277	62,1348	60,7128	59,3347
70	112,3167	107,8079	104,2148	100,4251	95,0231	90,5313	85,5270	82,2553	79,7147	77,5766	75,6893	73,9677	72,3583	70,8236	69,3345
80	124,8389	120,1018	116,3209	112,3288	106,6285	101,8795	96,5782	93,1058	90,4053	88,1303	86,1197	84,2840	82,5663	80,9266	79,3343
90	137,2082	132,2554	128,2987	124,1162	118,1359	113,1452	107,5650	103,9040	101,0537	98,6499	96,5238	94,5809	92,7614	91,0234	89,3342
100	149,4488	144,2925	140,1697	135,8069	129,5613	124,3421	118,4980	114,6588	111,6667	109,1412	106,9058	104,8615	102,9459	101,1149	99,3341
120	173,6184	168,0814	163,6485	158,9500	152,2113	146,5673	140,2326	136,0620	132,8063	130,0546	127,6159	125,3833	123,2890	121,2850	119,3340
140	197,4498	191,5653	186,8465	181,8405	174,6478	168,6130	161,8270	157,3517	153,8537	150,8941	148,2686	145,8629	143,6043	141,4413	139,3339
160	221,0197	214,8081	209,8238	204,5300	196,9152	190,5164	183,3106	178,5517	174,8283	171,6752	168,8759	166,3092	163,8977	161,5868	159,3338
180	244,3723	237,8548	232,6198	227,0563	219,0442	212,3039	204,7036	199,6786	195,7434	192,4086	189,4462	186,7282	184,1732	181,7234	179,3338
200	267,5388	260,7350	255,2638	249,4452	241,0578	233,9942	226,0210	220,7441	216,6088	213,1022	209,9854	207,1244	204,4337	201,8526	199,3337
250	324,8306	317,3609	311,3460	304,9393	295,6885	287,8815	279,0504	273,1944	268,5987	264,6970	261,2253	258,0355	255,0327	252,1497	249,3337
300	381,4239	373,3509	366,8439	359,9064	349,8745	341,3951	331,7885	325,4090	320,3971	316,1383	312,3460	308,8589	305,5741	302,4182	299,3336
500	603,4458	593,3580	585,2060	576,4931	563,8514	553,1269	540,9303	532,8028	526,4014	520,9505	516,0874	511,6081	507,3816	503,3147	499,3335
600	712,7726	701,8322	692,9809	683,5155	669,7690	658,0936	644,8004	635,9329	628,8157	622,9876	617,6713	612,7718	608,1468	603,6942	599,3335

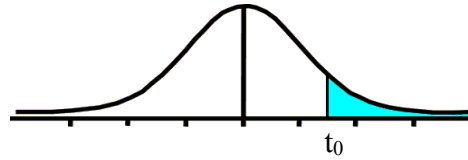
TABLA 3-Distribución Chi Cuadrado χ^2 . (Continuación)

v/p	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	0,975	0,99	0,995	0,9975	0,999
1	0,3573	0,2750	0,2059	0,1485	0,1015	0,0642	0,0358	0,0158	0,0039	0,0010	0,0002	0,0000	0,0000	0,0000
2	1,1957	1,0217	0,8616	0,7133	0,5754	0,4463	0,3250	0,2107	0,1026	0,0506	0,0201	0,0100	0,0050	0,0020
3	2,1095	1,8692	1,6416	1,4237	1,2125	1,0052	0,7978	0,5844	0,3518	0,2158	0,1148	0,0717	0,0449	0,0243
4	3,0469	2,7528	2,4701	2,1947	1,9226	1,6488	1,3665	1,0636	0,7107	0,4844	0,2971	0,2070	0,1449	0,0908
5	3,9959	3,6555	3,3251	2,9999	2,6746	2,3425	1,9938	1,6103	1,1455	0,8312	0,5543	0,4118	0,3075	0,2102
6	4,9519	4,5702	4,1973	3,8276	3,4546	3,0701	2,6613	2,2041	1,6354	1,2373	0,8721	0,6757	0,5266	0,3810
7	5,9125	5,4932	5,0816	4,6713	4,2549	3,8223	3,3583	2,8331	2,1673	1,6899	1,2390	0,9893	0,7945	0,5985
8	6,8766	6,4226	5,9753	5,5274	5,0706	4,5936	4,0782	3,4895	2,7326	2,1797	1,6465	1,3444	1,1042	0,8571
9	7,8434	7,3570	6,8763	6,3933	5,8988	5,3801	4,8165	4,1682	3,3251	2,7004	2,0879	1,7349	1,4501	1,1519
10	8,8124	8,2955	7,7832	7,2672	6,7372	6,1791	5,5701	4,8652	3,9403	3,2470	2,5582	2,1558	1,8274	1,4787
11	9,7831	9,2373	8,6952	8,1479	7,5841	6,9887	6,3364	5,5778	4,5748	3,8157	3,0535	2,6032	2,2321	1,8338
12	10,7553	10,1820	9,6115	9,0343	8,4384	7,8073	7,1138	6,3038	5,2260	4,4038	3,5706	3,0738	2,6612	2,2141
13	11,7288	11,1291	10,5315	9,9257	9,2991	8,6339	7,9008	7,0415	5,8919	5,0087	4,1069	3,5650	3,1118	2,6172
14	12,7034	12,0785	11,4548	10,8215	10,1653	9,4673	8,6963	7,7895	6,5706	5,6287	4,6604	4,0747	3,5820	3,0407
15	13,6790	13,0298	12,3809	11,7212	11,0365	10,3070	9,4993	8,5468	7,2609	6,2621	5,2294	4,6009	4,0697	3,4825
16	14,6555	13,9827	13,3096	12,6243	11,9122	11,1521	10,3090	9,3122	7,9616	6,9077	5,8122	5,1422	4,5734	3,9417
17	15,6328	14,9373	14,2406	13,5307	12,7919	12,0023	11,1249	10,0852	8,6718	7,5642	6,4077	5,6973	5,0916	4,4162
18	16,6108	15,8932	15,1738	14,4399	13,6753	12,8570	11,9462	10,8649	9,3904	8,2307	7,0149	6,2648	5,6234	4,9048
19	17,5894	16,8504	16,1089	15,3517	14,5620	13,7158	12,7727	11,6509	10,1170	8,9065	7,6327	6,8439	6,1673	5,4067
20	18,5687	17,8088	17,0458	16,2659	15,4518	14,5784	13,6039	12,4426	10,8508	9,5908	8,2604	7,4338	6,7228	5,9210
21	19,5485	18,7683	17,9843	17,1823	16,3444	15,4446	14,4393	13,2396	11,5913	10,2829	8,8972	8,0336	7,2889	6,4467
22	20,5288	19,7288	18,9243	18,1007	17,2396	16,3140	15,2787	14,0415	12,3380	10,9823	9,5425	8,6427	7,8648	6,9829
23	21,5095	20,6902	19,8657	19,0211	18,1373	17,1865	16,1219	14,8480	13,0905	11,6885	10,1957	9,2604	8,4503	7,5291
24	22,4908	21,6525	20,8084	19,9432	19,0373	18,0618	16,9686	15,6587	13,8484	12,4011	10,8563	9,8862	9,0441	8,0847
25	23,4724	22,6156	21,7524	20,8670	19,9393	18,9397	17,8184	16,4734	14,6114	13,1197	11,5240	10,5196	9,6462	8,6494
26	24,4544	23,5794	22,6975	21,7924	20,8434	19,8202	18,6714	17,2919	15,3792	13,8439	12,1982	11,1602	10,2561	9,2222
27	25,4367	24,5440	23,6437	22,7192	21,7494	20,7030	19,5272	18,1139	16,1514	14,5734	12,8785	11,8077	10,8733	9,8029
28	26,4195	25,5092	24,5909	23,6475	22,6572	21,5880	20,3857	18,9392	16,9279	15,3079	13,5647	12,4613	11,4973	10,3907
29	27,4025	26,4751	25,5391	24,5770	23,5666	22,4751	21,2468	19,7677	17,7084	16,0471	14,2564	13,1211	12,1278	10,9861

TABLA 3-Distribución Chi Cuadrado χ^2 . (Continuación)

v/p	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	0,975	0,99	0,995	0,9975	0,999
30	28,3858	27,4416	26,4881	25,5078	24,4776	23,3641	22,1103	20,5992	18,4927	16,7908	14,9535	13,7867	12,7646	11,5876
31	29,3694	28,4087	27,4381	26,4397	25,3901	24,2551	22,9762	21,4336	19,2806	17,5387	15,6555	14,4577	13,4073	12,1961
32	30,3533	29,3763	28,3889	27,3728	26,3041	25,1478	23,8442	22,2706	20,0719	18,2908	16,3622	15,1340	14,0555	12,8104
33	31,3375	30,3444	29,3405	28,3069	27,2194	26,0422	24,7143	23,1102	20,8665	19,0467	17,0735	15,8152	14,7092	13,4312
34	32,3219	31,3130	30,2928	29,2421	28,1361	26,9383	25,5864	23,9522	21,6643	19,8062	17,7891	16,5013	15,3679	14,0568
35	33,3065	32,2821	31,2458	30,1782	29,0540	27,8359	26,4604	24,7966	22,4650	20,5694	18,5089	17,1917	16,0315	14,6881
36	34,2913	33,2517	32,1995	31,1152	29,9730	28,7350	27,3363	25,6433	23,2686	21,3359	19,2326	17,8868	16,7000	15,3243
37	35,2764	34,2216	33,1539	32,0532	30,8933	29,6355	28,2138	26,4921	24,0749	22,1056	19,9603	18,5859	17,3730	15,9652
38	36,2617	35,1920	34,1089	32,9919	31,8146	30,5373	29,0931	27,3430	24,8839	22,8785	20,6914	19,2888	18,0501	16,6109
39	37,2472	36,1628	35,0645	33,9315	32,7369	31,4405	29,9739	28,1958	25,6954	23,6543	21,4261	19,9958	18,7318	17,2612
40	38,2328	37,1340	36,0207	34,8719	33,6603	32,3449	30,8563	29,0505	26,5093	24,4331	22,1642	20,7066	19,4171	17,9166
45	43,1638	41,9950	40,8095	39,5847	38,2910	36,8844	35,2895	33,3504	30,6123	28,3662	25,9012	24,3110	22,8994	21,2509
50	48,0986	46,8638	45,6100	44,3133	42,9421	41,4492	39,7539	37,6886	34,7642	32,3574	29,7067	27,9908	26,4636	24,6736
55	53,0367	51,7391	50,4204	49,0554	47,6105	46,0356	44,2448	42,0596	38,9581	36,3981	33,5705	31,7349	30,0974	28,1731
60	57,9775	56,6200	55,2394	53,8091	52,2938	50,6406	48,7587	46,4589	43,1880	40,4817	37,4848	35,5344	33,7909	31,7381
70	67,8664	66,3961	64,8990	63,3460	61,6983	59,8978	57,8443	55,3289	51,7393	48,7575	45,4417	43,2753	41,3323	39,0358
80	77,7631	76,1879	74,5825	72,9153	71,1445	69,2070	66,9938	64,2778	60,3915	57,1532	53,5400	51,1719	49,0430	46,5197
90	87,6661	85,9925	84,2854	82,5111	80,6247	78,5584	76,1954	73,2911	69,1260	65,6466	61,7540	59,1963	56,8918	54,1559
100	97,5744	95,8078	94,0046	92,1290	90,1332	87,9453	85,4406	82,3581	77,9294	74,2219	70,0650	67,3275	64,8571	61,9182
120	117,4041	115,4646	113,4825	111,4186	109,2197	106,8056	104,0374	100,6236	95,7046	91,5726	86,9233	83,8517	81,0726	77,7555
140	137,2476	135,1491	133,0028	130,7657	128,3800	125,7580	122,7476	119,0293	113,6594	109,1368	104,0343	100,6547	97,5908	93,9253
160	157,1019	154,8555	152,5564	150,1583	147,5988	144,7834	141,5475	137,5457	131,7560	126,8700	121,3457	117,6791	114,3496	110,3592
180	176,9652	174,5799	172,1373	169,5879	166,8653	163,8682	160,4206	156,1526	149,9687	144,7413	138,8205	134,8843	131,3050	127,0114
200	196,8359	194,3193	191,7409	189,0486	186,1717	183,0028	179,3550	174,8353	168,2785	162,7280	156,4321	152,2408	148,4262	143,8420
250	246,5387	243,7202	240,8297	237,8085	234,5768	231,0128	226,9048	221,8059	214,3915	208,0978	200,9387	196,1604	191,8020	186,5537
300	296,2700	293,1786	290,0062	286,6878	283,1353	279,2143	274,6901	269,0679	260,8781	253,9122	245,9727	240,6631	235,8126	229,9620
500	495,3734	491,3709	487,2569	482,9462	478,3231	473,2099	467,2962	459,9261	449,1467	439,9360	429,3874	422,3034	415,8081	407,9458
600	594,9938	590,6057	586,0930	581,3623	576,2859	570,6681	564,1661	556,0560	544,1801	534,0185	522,3654	514,5285	507,3385	498,6219

Tabla t-Student



Grados de libertad	0.25	0.1	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.8165	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.7407	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7970
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.6800	1.3007	1.6794	2.0141	2.4121	2.6896
46	0.6799	1.3002	1.6787	2.0129	2.4102	2.6870
47	0.6797	1.2998	1.6779	2.0117	2.4083	2.6846
48	0.6796	1.2994	1.6772	2.0106	2.4066	2.6822
49	0.6795	1.2991	1.6766	2.0096	2.4049	2.6800

50	0.6794	1.2987	1.6759	2.0086	2.4033	2.6778
51	0.6793	1.2984	1.6753	2.0076	2.4017	2.6757
52	0.6792	1.2980	1.6747	2.0066	2.4002	2.6737
53	0.6791	1.2977	1.6741	2.0057	2.3988	2.6718
54	0.6791	1.2974	1.6736	2.0049	2.3974	2.6700
55	0.6790	1.2971	1.6730	2.0040	2.3961	2.6682
56	0.6789	1.2969	1.6725	2.0032	2.3948	2.6665
57	0.6788	1.2966	1.6720	2.0025	2.3936	2.6649
58	0.6787	1.2963	1.6716	2.0017	2.3924	2.6633
59	0.6787	1.2961	1.6711	2.0010	2.3912	2.6618
60	0.6786	1.2958	1.6706	2.0003	2.3901	2.6603
61	0.6785	1.2956	1.6702	1.9996	2.3890	2.6589
62	0.6785	1.2954	1.6698	1.9990	2.3880	2.6575
63	0.6784	1.2951	1.6694	1.9983	2.3870	2.6561
64	0.6783	1.2949	1.6690	1.9977	2.3860	2.6549
65	0.6783	1.2947	1.6686	1.9971	2.3851	2.6536
66	0.6782	1.2945	1.6683	1.9966	2.3842	2.6524
67	0.6782	1.2943	1.6679	1.9960	2.3833	2.6512
68	0.6781	1.2941	1.6676	1.9955	2.3824	2.6501
69	0.6781	1.2939	1.6672	1.9949	2.3816	2.6490
70	0.6780	1.2938	1.6669	1.9944	2.3808	2.6479
71	0.6780	1.2936	1.6666	1.9939	2.3800	2.6469
72	0.6779	1.2934	1.6663	1.9935	2.3793	2.6458
73	0.6779	1.2933	1.6660	1.9930	2.3785	2.6449
74	0.6778	1.2931	1.6657	1.9925	2.3778	2.6439
75	0.6778	1.2929	1.6654	1.9921	2.3771	2.6430
76	0.6777	1.2928	1.6652	1.9917	2.3764	2.6421
77	0.6777	1.2926	1.6649	1.9913	2.3758	2.6412
78	0.6776	1.2925	1.6646	1.9908	2.3751	2.6403
79	0.6776	1.2924	1.6644	1.9905	2.3745	2.6395
80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387
81	0.6775	1.2921	1.6639	1.9897	2.3733	2.6379
82	0.6775	1.2920	1.6636	1.9893	2.3727	2.6371
83	0.6775	1.2918	1.6634	1.9890	2.3721	2.6364
84	0.6774	1.2917	1.6632	1.9886	2.3716	2.6356
85	0.6774	1.2916	1.6630	1.9883	2.3710	2.6349
86	0.6774	1.2915	1.6628	1.9879	2.3705	2.6342
87	0.6773	1.2914	1.6626	1.9876	2.3700	2.6335
88	0.6773	1.2912	1.6624	1.9873	2.3695	2.6329
89	0.6773	1.2911	1.6622	1.9870	2.3690	2.6322
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316
91	0.6772	1.2909	1.6618	1.9864	2.3680	2.6309
92	0.6772	1.2908	1.6616	1.9861	2.3676	2.6303
93	0.6771	1.2907	1.6614	1.9858	2.3671	2.6297
94	0.6771	1.2906	1.6612	1.9855	2.3667	2.6291
95	0.6771	1.2905	1.6611	1.9852	2.3662	2.6286
96	0.6771	1.2904	1.6609	1.9850	2.3658	2.6280
97	0.6770	1.2903	1.6607	1.9847	2.3654	2.6275
98	0.6770	1.2903	1.6606	1.9845	2.3650	2.6269
99	0.6770	1.2902	1.6604	1.9842	2.3646	2.6264
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758