

PROYECTO DE GRADO

Presentado ante la ilustre UNIVERSIDAD DE LOS ANDES como requisito parcial para
obtener el Título de INGENIERO DE SISTEMAS

MODELOS ESTADÍSTICOS PARA PRONOSTICAR CRIPTOMONEDAS. CASO: BITCOIN, ETHEREUM Y LITECOIN.

Por

www.bdigital.ula.ve

Br. Keiswuer Emiro Chirinos Hermanadez

Tutor: Profa. Laura Daniela Castillo Paredes

Octubre del 2022



©2022 Universidad de Los Andes Mérida, Venezuela

C.C. Reconocimiento

Modelos Estadísticos para Pronosticar Criptomonedas. Caso: Bitcoin, Ethereum y Litecoin.

Br. Keiswuer Emiro Chirinos Hermanadez

Proyecto de Grado — Investigación de Operaciones, 113 páginas

Resumen: En el presente proyecto se hace una recopilación teórica sobre el funcionamiento, características, ventajas, desventajas y expectativas a futuro de las criptomonedas, con énfasis en el estudio del pronóstico. Para el estudio se analizó diversos modelos estadísticos de predicción, tomando como elección la regresión lineal y series temporales. Una vez expuestos los conceptos teóricos, se ejecutó varias funciones en R, que permitió hacer los cálculos con rapidez y confiabilidad de los modelos planteados, además se estudió las pruebas de hipótesis necesaria para garantizar el cumplimiento de los requerimientos teóricos. Finalmente, se aplicó la teoría de los modelos de predicción seleccionados y se obtuvo resultados confiables, en un lapso de 30 días. La regresión lineal, para un instante determinado es un modelo predictivo bastante efectivo, proporciona buenas predicciones considerando la gran volatilidad que presentan estos criptoactivos, sin embargo, las series temporales resultaron ser bastante efectivas para las criptomonedas seleccionadas.

Palabras clave: Criptomonedas, Modelos Predictivos, Bitcoin, Ethereum, Litecoin.

Índice

1	Introducción	1
1.1	Antecedentes	2
1.2	Planteamiento del Problema	3
1.3	Justificación	4
1.4	Formulación del Problema	4
1.5	Objetivos	4
1.5.1	Objetivo General	4
1.5.2	Objetivos Específicos	5
1.6	Alcance	5
1.7	Metodología	5
2	Definición y Conceptos Claves de las Criptomonedas	7
2.1	Historia de las Criptomonedas	7
2.2	Descripción de la Criptografía	9
2.3	Descripción de la Blockchain	10
2.4	Minería de Criptomonedas	13
2.5	Ventajas y Desventajas de las Criptomonedas	14
2.5.1	Ventajas	14
2.5.2	Desventajas	15
2.6	Características de las Criptomonedas	15
2.6.1	Libre Acceso	16
2.6.2	Descentralización	16
2.6.3	Anonimato	16

2.6.4	Transparencia	16
2.6.5	Volatilidad	17
2.6.6	Regulación	17
2.6.7	Irreversibilidad	17
2.6.8	Control Total de tu Dinero	17
2.6.9	Código Abierto	18
2.6.10	Integridad	18
2.7	Expectativa de las Criptomonedas en el Futuro	18
2.7.1	Adopción	19
2.7.2	Viabilidad	19
2.7.3	Versatilidad	19
2.8	Formas de Inversión en Criptomonedas	19
2.8.1	Análisis Fundamental	20
2.8.2	Análisis Técnico	20
2.9	Descripción de las Criptomonedas Seleccionadas	21
2.9.1	BITCOIN	21
2.9.2	ETHEREUM	22
2.9.3	LITECOIN	24
3	Descripción de los Modelos de Predicción	26
3.1	Introducción	26
3.2	Regresión Lineal	27
3.2.1	Regresión Lineal Simple	27
3.2.2	Criterio de los Mínimos Cuadrados	29
3.2.3	Pruebas de Hipótesis en la Regresión	30
3.2.4	Análisis de Varianza	31
3.2.5	Calidad del Ajuste	32
3.2.6	Análisis de Residuales	33
3.2.7	Verificación del Supuesto de Independencia	35
3.2.8	Predicción a Partir del Modelo	37
3.3	Series de Tiempo	38
3.3.1	Componentes de una Serie Temporal	38

3.3.2	Clasificación Descriptiva de las Series Temporales	39
3.3.3	Procesos Estocásticos	40
3.3.4	Ruido Blanco (White Noise)	41
3.3.5	Camino Aleatorio (Random Walks)	41
3.3.6	Función de Autocovarianza (ACV)	42
3.3.7	Función de Autocorrelación	43
3.3.8	Proceso Lineal Estacionario	46
3.3.9	Proceso Lineal No Estacionario	52
4	Modelización de los Modelos Planteados	64
4.1	Regresión Lineal Simple	64
4.1.1	Bitcoin	64
4.1.2	Ethereum	72
4.1.3	Litecoin	76
4.2	Series de Tiempo	80
4.2.1	Bitcoin	80
4.2.2	Ethereum	94
4.2.3	Litecoin	101
5	Conclusiones y Recomendaciones	109
	Bibliografía	111

Capítulo 1

Introducción

Las innovaciones tecnológicas cada día nos impulsan a nuevos avances, uno de ellos es el comercio electrónico que tiene el potencial de intercambiar bienes, servicios y activos a través del internet; motivando a que las personas adopten este medio. Unas de las alternativas más populares son las criptomonedas por tener la ventaja de no depender ni ser controlada por una institución financiera, permitiéndoles a su vez, que los usuarios tengan un control total sobre sus divisas en criptomonedas.

Las criptomonedas actualmente se están posicionando en los primeros puestos de intercambio digital por su elevado precio. Pronosticar el comportamiento de sus precios será un aporte útil para todos sus usuarios e inversionista que hacen uso de este cryptoactivo de intercambio digital, ya que cumple con las características que tiene el dinero, divisibilidad, durabilidad, fungibilidad y escasez. Son descentralizadas y todas sus transacciones son irreversibles y validadas por su propio sistema bajo la arquitectura de la Blockchain.

Existen países que han adoptado las criptomonedas como un medio corriente de transacción para compra y venta de bienes y servicios, permitiéndoles incrementar tanto de usuarios como en sus precios, sin embargo, existen varias problemáticas sobre este cryptoactivo digital, unas de ellas son; el desconocimiento que tienen las personas sobre sus mecanismos de operación, los riesgos que están expuesto los inversionistas

por su alta volatilidad en los precios, esto es, cuando ocurre un suceso pequeño, puede que su valor se dispare o disminuya radicalmente.

Este proyecto pretende mostrar, que empleando modelos estadísticos se pueden ajustar los datos, para estimar valores futuros de los precios de las criptomonedas como: Bitcoin, Ethereum Y Litecoin.

1.1 Antecedentes

Actualmente existen un conjunto de estudio en las cuales se han aplicado diversos modelos y análisis estadísticos para pronosticar las criptomonedas. A continuación, se hacen referencias algunas de estas investigaciones

MODELOS DE ANÁLISIS DE CRIPTOMONEDAS BASADO EN APRENDIZAJE AUTOMÁTICO Y SERIES TEMPORALES [20], Miguel Andrés Porro Vargas, Julio del 2018. Desarrollo un modelo de análisis basado en aprendizaje automático y series temporales para realizar pronósticos de precios. Resultando que los modelos ARIMA son altamente efectivos para la predicción de los precios, mostrando que es posible predecir con altos niveles de exactitud el comportamiento futuro del precio de las criptomonedas, mediante el uso de estos modelos de análisis.

LAS CRIPTOMONEDAS EN EL SISTEMA FINANCIERO INTERNACIONAL. UN ANÁLISIS PREDICTIVO MEDIANTE TÉCNICAS DE MACHINE LEARNING (LSTM) [21], Arnau Castillo Mur, junio 2020. Se analizó la capacidad predictiva del precio de cierre de modelos de Machine Learning como son las redes LSTM y de modelos estadísticos como el modelo ARIMA. Resulto que la capacidad predictiva de la evolución del precio destaco la red LSTM por encima del modelo ARIMA, consiguiendo una accuracy superior al 50% en casi todas las criptomonedas estudiadas.

ANTICIPANDO LOS PRECIOS DE LA CRIPTOMONEDA USANDO EL APRENDIZAJE AUTOMÁTICO, LAURA ALESSANDRETTI, Abeer ElBahrawy,

Luca María Aiello [22], y Andrea Baronchelli, 17 de octubre de 2018. El artículo estudia el rendimiento de tres modelos de pronósticos en los precios diarios de las criptomonedas. Dos de ellas se basaron en los árboles de decisiones de aumento del gradiente y uno se basó en redes neuronales recurrentes a largo y corto plazo, donde la predicción se basa en los precios anteriores de la moneda. Las métricas que caracterizan las criptomonedas que se utilizan a lo largo del estudio, los algoritmos de pronóstico y la evaluación de métricas. Resultaron que los métodos basados en arboles de decisión de aumento de gradiente funcionaron mejor cuando las predicciones se basaron en ventanas de corto plazo de 5/10 días, lo que sugiere que explotan bien en su mayoría dependencias de corto plazo. En cambio, las redes neuronales recurrentes funcionaron mejor cuando las predicciones se basaron en 50 días de datos, ya que también pueden capturar dependencias a largo plazo y son muy estables frente a la volatilidad de los precios.

1.2 Planteamiento del Problema

A nivel mundial las criptomonedas han tomado gran popularidad estos últimos años con la aparición del covid-19. Se ha elevado la adquisición de este criptoactivo por parte de las personas, ya que les permite hacer compra y venta de bienes y servicios sin tener que salir de sus casas, brindándoles seguridad en cada una de sus transacciones gracias a la tecnología de la Blockchain.

Desde el ámbito local Venezuela es uno de los países que han adoptado las criptomonedas como opción de inversión por la alta inflación que atraviesa el país, por esta razón el objetivo del proyecto es determinar los modelos estadísticos que pronostiquen mejor el comportamiento de las criptomonedas como Bitcoin, Ethereum y Litecoin, con el fin de observar expectativas a futuro y lograr disminuir los riesgos al momento de su inversión.

1.3 Justificación

Las criptomonedas se pueden considerar como parte fundamental a la independencia de las instituciones gubernamentales y banco centrales que pueden estar expuestos a corrupción, fraude o fallas en sus sistemas, permitiendo que los usuarios adopten de forma progresiva el uso de este criptoactivo digital como medio de inversión. Sin embargo, tienen la problemática de ser altamente volátiles, por esta razón el objetivo de este proyecto es pronosticar las criptomonedas, y así, brindar una mayor seguridad a los usuarios que quieran invertir en ellas.

La realización de este proyecto tiene como fin, aportar aspectos teóricos de la descripción del funcionamiento de las criptomonedas, al igual que los modelos seleccionados de predicción. Contará de diversas investigaciones y aplicaciones de herramientas para el ajuste de los datos. Así mismo, se ejecuta las predicciones y comparaciones entre los modelos para disminuir las complejidades en el comportamiento de las criptomonedas, y optar por soluciones correctas en el desarrollo de expectativas a futuro.

1.4 Formulación del Problema

Las consideraciones expuestas nos llevan a plantear la siguiente interrogantes:

- ¿Qué modelo estadístico replican mejor el comportamiento de las criptomonedas?
- ¿Qué modelo estadístico predicen mejor el comportamiento de las criptomonedas?

1.5 Objetivos

1.5.1 Objetivo General

Determinar los modelos estadísticos que predigan mejor el comportamiento de las criptomonedas: Bitcoin, Ethereum y Litecoin.

1.5.2 Objetivos Específicos

- Describir el funcionamiento de las criptomonedas.
- Describir los principales modelos estadísticos de predicción seleccionados.
- Ejecutar técnicas y herramientas que permitan ajustar los datos a los modelos seleccionados.
- Predecir el comportamiento de las criptomonedas. Caso: Bitcoin, Ethereum y Litecoin.
- Construir herramientas que permitan hacer comparaciones de los modelos ejecutados.
- Seleccionar el modelo estadístico que mejor prediga el comportamiento de las criptomonedas. Caso: Bitcoin, Ethereum y Litecoin.

1.6 Alcance

El proyecto tiene como alcance predecir el comportamiento de los precios de las criptomonedas, tomando como estudio los datos del Bitcoin, Ethereum y Litecoin de la página CoinMarketcap. Utilizaremos diferentes estrategias y herramientas disponibles para su ejecución, con el fin de elegir un modelo adecuado para obtener pronósticos acertados que sean de gran utilidad para todos los usuarios. Será un aporte fundamental como modelo piloto de investigaciones futuras, por sus infinitas aplicaciones.

1.7 Metodología

La metodología del estudio se iniciará describiendo el funcionamiento de las criptomonedas, tomando como elección al Bitcoin, Ethereum y Litecoin, fueron seleccionadas debido a su porcentaje de capitalización de mercado, siendo las más rentables y preferidas por los usuarios. Los datos serán recaudados de la página CoinMarketcap [4]. Así mismo se describirán cada uno de los modelos estadísticos que

se implementarán para la ejecución de las predicciones de las criptomonedas.

Se ejecutarán cada uno de los modelos estadísticos descritos de predicción, utilizando el lenguaje de programación R, bajo el entorno de desarrollo integrado RStudio, ejecutando las diversas herramientas a los datos para abordar las predicciones.

Posteriormente se analizarán los resultados de cada uno de los modelos, realizándoles comparaciones para tomar índices que surgen a partir de ellos, y así, garantizar un buen pronóstico de las criptomonedas estudiadas.

www.bdigital.ula.ve

C.C. Reconocimiento

Capítulo 2

Definición y Conceptos Claves de las Criptomonedas

2.1 Historia de las Criptomonedas

Las criptomonedas no son las primeras monedas digitales, ya existían a través de tarjetas de débito, crédito, transferencias, etc. Surgieron en los Estados Unidos durante la década de 1920 cuando Western Union comenzó a entregar a un grupo selecto de sus clientes una placa de metal que le permitiría identificar y definir sus pagos; posteriormente otras compañías como General Petroleum Corporation y diferentes cadenas hoteleras, siguieron esa idea tan brillante y ofrecieron algo similar a sus clientes. No obstante, el cambio significativo fue en 1960, cuando IBM (Internacional Business Machine Corporation) inventó la banda magnética para ser utilizada en el transporte público en Londres, Inglaterra. En ese mismo año, en Estados Unidos se emitieron las primeras tarjetas de crédito en las principales cadenas como Sears, ATyT, Hoteles Hilton y Asociación de Líneas Áreas, poniendo de manera masiva la circulación del dinero digital o criptomonedas.

Las primeras ideas a cerca de las criptomonedas, tal como se concibe actualmente el término, comenzaron a finales de 1980 y a principio de 1990, donde surge un movimiento denominado Cypherpunks, que abogaba por el uso de la criptografía como

herramienta de cambio social y político para lograr la libertad, privacidad y respeto sobre los servicios cibernéticos que comenzaban a expandirse en gran parte de Estados Unidos. No obstante el concepto o idea de criptomoneda fue descrita por primera vez por Wei Dai en 1998, donde propuso la idea de crear un nuevo tipo de dinero descentralizado que usara la criptografía como medio de control.

Pero no es hasta 2008, en medio de la crisis financiera de los Estados Unidos que devaluó el dólar, con los desafíos económicos que afectaron al mundo entero, y los bancos se encontraban en problemas debido al bajo valor de la moneda y al recorte en las tasas de interés, cuando el desarrollador de pseudónimo Satoshi Nakamoto publica en ese mismo año por primera vez su artículo sobre la tecnología de Bitcoin detallando el sistema “peer-to-peer” que ejecuta las transacciones de persona a persona, basada en un esquema de registro público global completamente descentralizado y con características criptográficas. (Bitcoin un sistema de efectivo electrónico usuario a usuario)[3].

Meses después, en 2009, el mismo Satoshi Nakamoto proporcionó el software para realizar las transacciones, y es como nace la primera criptomoneda llamada Bitcoin, que es una moneda digital completamente encriptada y descentralizada, abierta a todos sin ningún banco, gobierno o intermediario que la controle, gracias a la tecnología Blockchain. En 2010 con las primeras criptomonedas creadas y en circulación comienzan a realizar transacciones y actividades con bróker MtGox que rápidamente se convierte en el centro de negociación de referencia para las operaciones del Bitcoin. Trágicamente esta entidad en 2011 presenta dificultades informáticas generando pérdidas a inversionistas, y genera una caída en la cotización del Bitcoin llegando a un centavo estadounidense; en octubre de 2013 en Vancouver se pone en funcionamiento el primer cajero automático para la adquisición del mismo con moneda fiduciaria; en febrero de 2014 bróker MtGox vuelve a sufrir nuevos ataques y se declara en bancarrota al reportar la pérdida de 850.000 Bitcoin; posteriormente en el año 2016, Japón establece el Bitcoin como una moneda más para el uso de transacciones de compra y venta de bienes y servicios.

Ante todo el Bitcoin ha sido la madre de todas las criptomonedas, ya que a partir de ella han surgido innumerables criptomonedas que han adoptado su sistema, y siempre se ha mantenido como la primera con mayor capital de mercado de todas las criptomonedas que existen en el mundo. En la actualidad existen más de diez mil criptomonedas en todo el mundo y subiendo.

2.2 Descripción de la Criptografía

Es una técnica empleada para proteger datos y documentos. Su funcionamiento se basa a través de cifras y códigos que circulan en redes locales o en internet, se han utilizado desde hace miles de años para ocultar información que solo fueran descifradas por las personas que conocieran su significado. En la evolución computacional fue ampliamente modificada para crear algoritmos matemáticos que permitieran codificar mensaje, manteniendo la seguridad de los usuarios tanto del remitente como el destinatario, esto sucede a través del uso de una o más llaves. La llave es una secuencia de caracteres que puede contener letras, dígitos y símbolos (como una contraseña), y es convertida en un número utilizada por los métodos de criptografía para codificar y decodificar mensajes.

Las claves criptográficas pueden ser básicamente de dos tipos:

- **Simétricas:** Es la utilización de determinados algoritmos para descifrar y encriptar (ocultar) documentos. Son grupos de algoritmos distintos que se relacionan unos con otros para mantener la conexión confidencial de la información.
- **Asimétricas:** Según (Tecnología Informática)[18] es una fórmula matemática que utiliza dos llaves, una pública y la otra privada. La llave pública es aquella a la que cualquier persona puede tener acceso, mientras que la llave privada, es sólo la persona que la recibe es capaz de descifrar.

2.3 Descripción de la Blockchain

La Blockchain fue creada en 1991 pero no fue utilizada hasta 2009, donde Satoshi Nakamoto da a conocerlo a través de la criptomoneda Bitcoin como la primera plataforma Blockchain, con la idea de realizar transacciones de usuario a usuario sin tener que depender de instituciones gubernamentales, bancos o intermediarios gracias al sistema “peer-to-peer”, que es una red donde cada nodo se comporta de la misma manera, es decir cada uno actúa tanto de cliente como de servidor interconectados entre sí. Esto permite una serie de características muy útiles al momento de intercambiar información de manera descentralizada, además de esta forma logra que ningún nodo sea imprescindible para el funcionamiento de la red.

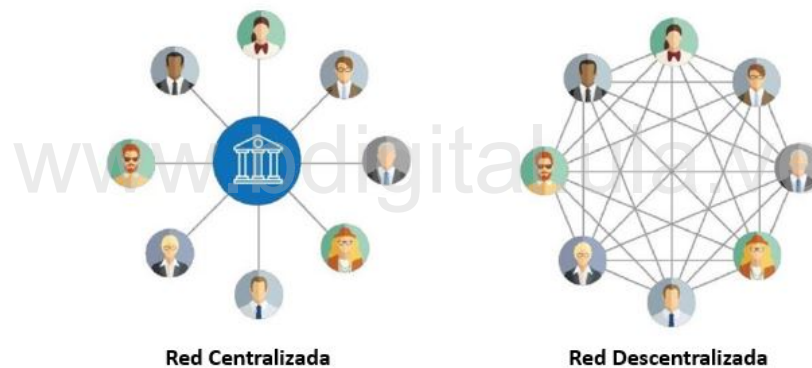


Figura 2.1: Diferencia del Funcionamiento de Redes

Fuente: Elaboración Propia.

La tecnología Blockchain funciona como una base de datos distribuida realizada por sus usuarios, es una articulación de tecnologías estructuradas en un sistema naturalmente encriptada, proporcionándoles a los usuarios involucrados la protección de sus identidades y de los datos de sus transacciones gracias a su arquitectura llamada “cadenas de bloques”.

Cada cadena de bloque puede contener diferentes tipos de información, donde permite una base de datos distribuida que registra bloques y los entrelaza para facilitar

la recuperación de la información y la verificación de que esta no ha sido cambiada.

Esto es gracias a los apuntadores Hash, que se trata del número de identificación del bloque, el cual es único e irrepetible. Cada bloque tiene uno propio, que conecta el bloque actual con el anterior y así sucesivamente hasta llegar al bloque génesis creando una cadena.

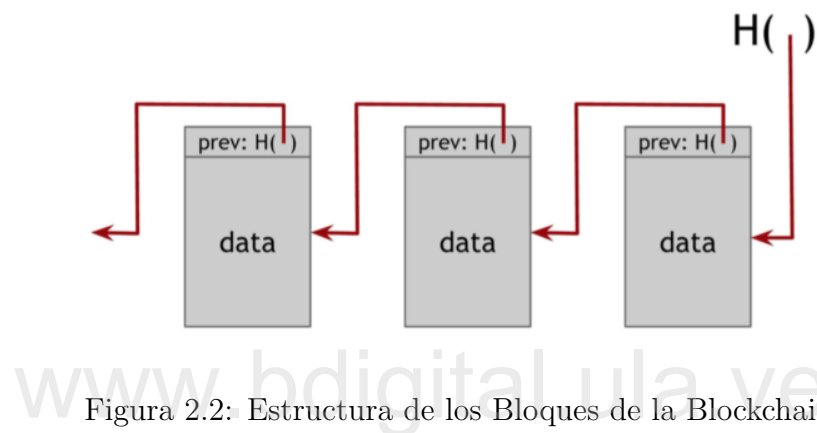


Figura 2.2: Estructura de los Bloques de la Blockchain

Fuente: [https:// Criptoinfo.com](https://Criptoinfo.com)

Este número se genera según el contenido del bloque, eso significa que si se cambia la información del bloque, automáticamente cambia el Hash y no permiten que validen la transacción; no es que exista una única base de datos, sino que cada usuario tiene una copia de los datos, y si trata de cambiar o alterar la información de su copia, la comunidad lo sabe y su versión de la base de datos es anulada y queda sin efecto, de modo que para completar el bloque actual de información tienen que tenerse en cuenta las transacciones de información producidas en bloques anteriores; esto provoca que la información sea imposible de borrar o modificar por ningún tipo de usuario, ya que la seguridad y la certificación de los documentos de Blockchain es suministrada por todos los usuarios de la red que conforman la cadena de bloques. Esta es la razón que ha impulsado a las industrias financieras y bancarias a dar señales de interés por la tecnología Blockchain.

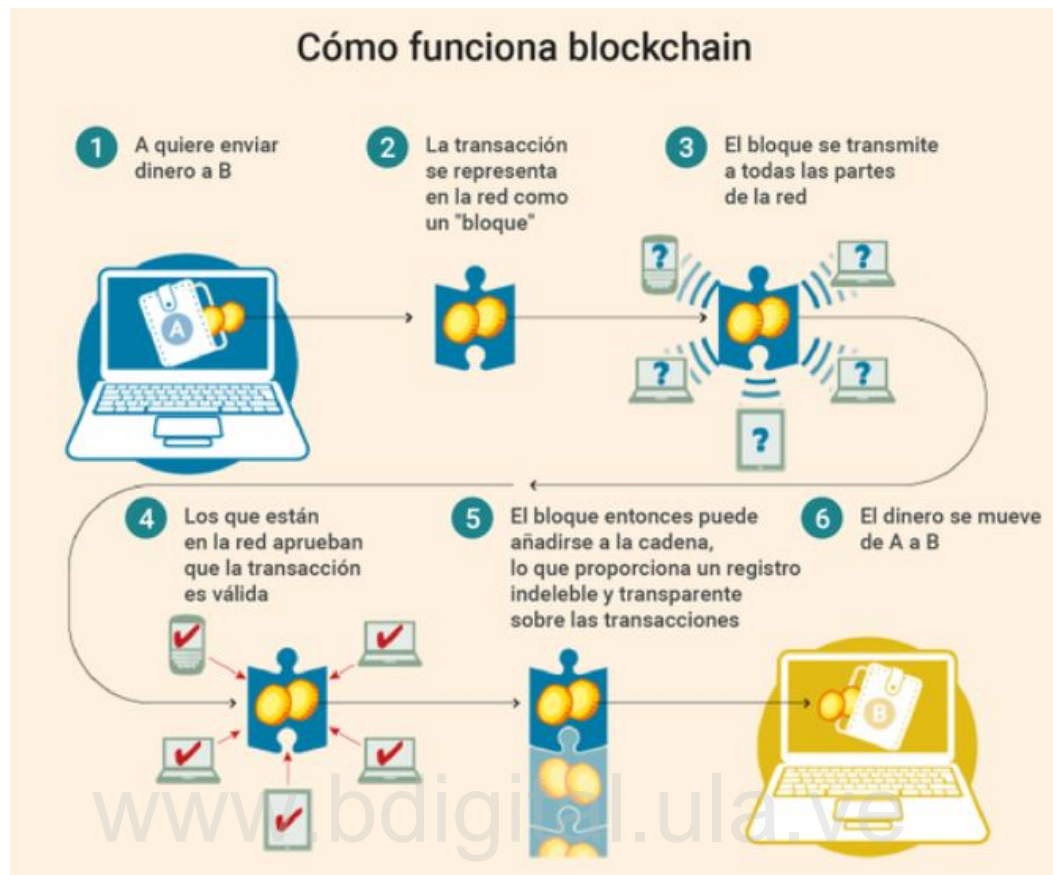


Figura 2.3: Funcionamiento de la Blockchain

Fuente: ft.com

La tecnología Blockchain surge junto al Bitcoin, pero al pasar los años ha tenido múltiples aplicaciones, no solo en el dinero digital si no en revoluciones, como los contratos inteligentes desarrollados en Ethereum, que permite que los programas actúen de forma automática e inteligente; por eso están siendo utilizadas por sectores empresariales y hasta por desarrolladores del Internet de las Cosas (IoT, por sus siglas en ingles).

Las aplicaciones de este sistema son muy diversas, ya que puede usarse para firmar contratos, votar en elecciones, guardar registros médicos y bancarios y muchas otras utilidades que todavía están por descubrir. Blockchain es más que una simple base de datos, es un sistema de almacenaje de información fuera del sistema convencional, una

herramienta para crear una sociedad más equitativa, más transparente y más veraz.

2.4 Minería de Criptomonedas

La minería de las criptomonedas, es el aporte computacional que cada minero le hace a la red P2P para mantenerla segura y anónima, logrando así confirmar todas las transacciones entre los usuarios. Un minero es un usuario que con su ordenador decide prestar la capacidad del mismo para que trabaje recolectando información encriptada, descifrándolas a fin de que sean escritas en una cadena de bloques (Blockchain). Para resolver un bloque se necesita de mucho aporte computacional, debido las complejidades de muchas operaciones matemáticas; este aporte computacional es medido en MGHZ, recibiendo a cambio una cantidad determinada de criptodivisas de la moneda que se está minando. La recompensa se hace de una cantidad u otra según el tiempo que tarden los mineros en descifrar la transacción; el primer minero que logre resolver la transacción es recompensado con mayor beneficio que sus competidores.

Para la creación de nuevos bloques se utiliza un software que conecta el ordenador del usuario con los ordenadores de los demás mineros de todo el mundo, para que trabajen en la búsqueda de un algoritmo que corresponda con las transacciones realizadas entre criptomonedas. Una vez recibida la confirmación, el bloque es creado, y el creador es recompensado con una cantidad de monedas, es así como se generan nuevas criptomonedas de las monedas minables. Hay muchos usuarios que han optado por ser parte de la minería de las criptomonedas, pero cada vez que se generan nuevos bloques la complejidad aumenta, por lo que hoy en día se necesitan de equipos adecuados para lograr resolver las complejidades de los bloques.

Existen diversas formas de minar criptomonedas, algunas de ellas son: Piscina de Minería (Pool Mining), que son grupos de usuarios que se asocian para acumular mayor poder computacional; se debe instalar un software, y la recompensa es suministrada en función de la capacidad de procesamiento que aporta a la piscina. Otra forma es el Minado en la Nube (Cloud Mining), que es un proceso de minado mediante un centro

de datos remoto con capacidad de procesamiento compartido, contratado a través de una empresa de minería en la nube.

2.5 Ventajas y Desventajas de las Criptomonedas

Los últimos años las criptomonedas han sido una innovación en el mundo económico, revolucionando el sector tecnológico gracias a la tecnología Blockchain. Son muchas las personas al rededor del mundo que utilizan las criptomonedas de forma normal, tanto para inversión como para realizar compra-venta de bienes y servicios; sin embargo hay una gran parte de la población que desconoce de sus beneficios y riesgos, es por ello que nos plantearemos dar a conocer algunas ventajas y desventajas de las criptomonedas (Cripto fácil)[5].

2.5.1 Ventajas

- Los pagos se realizan de forma instantánea y sin intermediarios gracias a la red P2P, que permite que las transacciones sean de usuario a usuario.
- Realizar los pagos con criptomonedas es cada día más sencillo, basta tan solo escanear el QR de la persona a la cual se le generará el pago e indicar la cantidad deseada, sin importar que nos encontremos en un país ajeno al registro de nuestra cartera, basta solo con un móvil o una computadora con internet.
- La adquisición de las criptomonedas se hace a través de un monedero; en resumen es un software que funciona como billetera o cartera virtual encriptada, otorgando seguridad total para guardar, enviar y recibir cualquier criptomoneda. Además los cobros porcentuales por cada transacción son mínimos.
- Las criptomonedas tienen un sistema criptográfico extremadamente seguro, hasta tal punto que ninguna ha presentado fallas en su seguridad. Es uno de los beneficios más fuertes gracias al sistema de encriptación en cada movimiento, debido a que existe una comunidad amplia que corrige hasta las irregularidades más mínimas.

- Los usuarios poseen una gran privacidad, ya que sus nombres, dirección, empleo o cualquier dato no es revelado, lo cual si es requisito para los bancos o instituciones; es por eso que su descentralización no permite que ningún país o institución pueda llegar a controlarlas.

2.5.2 Desventajas

- Es necesario contar con internet para hacer uso de ellas; sin este requisito es imposible acceder o hacer cualquier movimiento de las criptomonedas.
- Las criptomonedas requieren de una aceptación significativa de las personas para lograr que su uso sea generalizado.
- Actualmente las computadoras cuánticas no han salido al mercado, pero representan un riesgo a la seguridad de las criptomonedas. Estos son ordenadores que poseen alto poder de procesamiento, que hacen vulnerable el sistema criptográfico de las criptomonedas.
- Las criptomonedas presentan una gran volatilidad en sus precios, ya que se rigen en el fundamento de oferta y demanda, por lo tanto permite que pequeños sucesos aumenten o disminuyan el precio, esto es debido a que no existe una institución reconocida detrás; sin embargo, hay países que poseen reservas de divisas para controlar que su precio no decaiga.
- Existen muchos países que han prohibido algunas criptomonedas, evitando que su uso se fomente.
- Se debe contar con una copia de seguridad del wallet o cartera, de manera que las criptomonedas solo existen de forma digital, y se corre el riesgo de perder todo el dinero.

2.6 Características de las Criptomonedas

Las características de las criptomonedas son la parte fundamental para adquirirla como medio de pago o inversión. Existen muchas personas que aún desconocen

completamente sus características, es por ello que daremos a conocer una serie de características comunes de todas las criptomonedas.

2.6.1 Libre Acceso

Al día de hoy, el libre acceso que se tiene para ser uso de las criptomonedas, es muy importante para el impulso del mercado financiero, permitiendo contribuir a elevar los precios de las criptomonedas.

2.6.2 Descentralización

La descentralización de las criptomonedas permite que no sean reguladas por gobiernos o instituciones financieras, por lo tanto no están sujetas a ninguna política o agendas de Bancos Centrales. Las regulaciones y movimientos son realizados por la comunidad de usuarios (Criptonoticias) [7].

2.6.3 Anonimato

Es necesario contar con un monedero virtual, en el que debemos introducir nuestros datos personales; no es por necesidad de control, si no para tener una copia de seguridad, ya que se trata de monedas digitales. Todas sus transacciones se hacen de forma anónima sin la necesidad de revelar ninguna información personal entre los usuarios, lo cual no ocurre en los sistemas financieros tradicionales (Qué es un wallet o monedero digital y criptomonedas)[15].

2.6.4 Transparencia

Todas las transacciones son registradas y guardadas en la Blockchain, permitiéndole a los usuarios revisar todas las transacciones de cualquier criptomoneda desde su aparición (Qué es la tecnología de contabilidad distribuida o blockchain)[14].

2.6.5 Volatilidad

La volatilidad de las criptomonedas son las fluctuaciones pronunciadas que se presentan en el precio. Las instituciones bancarias y financieras pensaron que el uso de las criptomonedas era una moda pasajera, ya que son demasiado volátiles para ser tomadas en serio, pero no resultó así, puesto que hay muchas personas y países que han aceptado su funcionamiento para compra y venta de bienes y servicios, aumentando en forma exponencial su circulación. Sin embargo la volatilidad podría anular la opción de ser adoptada como una moneda general en la sociedad, ya que su precio puede cambiar radicalmente por pequeños sucesos permitiendo que su valor se dispare o disminuya, generando riesgo para usuarios e inversionistas.

2.6.6 Regulación

El uso de las criptomonedas no está regulado por ninguna entidad. Según (La regulación de las criptomonedas)[9], el debate de la regulación de las criptomonedas cada vez está tomando más peso, y los gobiernos están preocupados por la seguridad del sistema financiero y el lavado de dinero. Quieren lograr proteger a los inversionistas que no tienen un conocimiento profundo de estos instrumentos financieros tan volátiles. De igual manera, pretenden frenar o disminuir el uso de las criptomonedas para evitar actividades ilegales.

2.6.7 Irreversibilidad

Las transacciones de las criptomonedas son irreversibles, una vez hecho el envío no hay forma de anularlo, deshacer una transacción es imposible, por ello es importante verificar que el pago se está haciendo de forma correcta para evitar pérdidas.

2.6.8 Control Total de tu Dinero

Las criptomonedas han capturado el interés de inversionistas, empresas y público en general por igual, hasta el punto que algunos usuarios la consideran un digno aspirante al trono que actualmente presenta el dinero fiduciario. Las criptomonedas tienen el

potencial de ahorrarle a usuarios y empresas de servicios financieros una cantidad significativa de tiempo y dinero al no contar con intermediarios en sus transacciones. Todo usuario tiene un control total de su dinero, gracias a la red P2P que le permite hacer cualquier transacción de usuario a usuario. La mayoría de criptomonedas no son emitidas por ningún país o estado, y por lo tanto no están sujetas a las mismas fluctuaciones geográficas del dinero fiduciario (¿Son las criptomonedas el nuevo dinero fiduciario?)[17].

2.6.9 Código Abierto

Casi todas las criptomonedas son desarrolladas en código abierto por su descentralización, permitiéndole a los usuarios en cualquier parte del mundo poder colaborar en su desarrollo de manera voluntaria, sin embargo para proponer cambiar un código se debe tener el consenso de ser aprobadas (Quienes están detrás del desarrollo de las criptomonedas)[16].

2.6.10 Integridad

La integridad financiera de las criptomonedas es gracias a la tecnología Blockchain. Se considera que es probablemente el desarrollo más importante de las criptomonedas, puesto que tienen el potencial de revolucionar la forma en que almacenamos los datos, concediendo una mayor seguridad y por tanto, una mejor calidad y confianza (Integridad de datos gracias a blockchain)[8]

2.7 Expectativa de las Criptomonedas en el Futuro

Es de pensar que el futuro de las criptomonedas es positivo, ya que su evolución cada día es más eficaz. Hasta ahora existen más de diez mil criptomonedas, y muchas de ellas se han enfocado en otros ámbitos, como los contratos inteligentes desarrollados por la criptomoneda Ethereum, permitiendo que a las personas les sea de interés, ayudando a expandir el uso de las criptomonedas a futuro.

2.7.1 Adopción

Para que una criptomoneda sea un éxito de negocios necesita una adopción global, pero esa aceptación por parte de las principales entidades financieras como medio de compra y venta de bienes y servicios, hasta ahora no es convincente para algunas personas y gobiernos. Sin embargo, puede que no sea así por mucho tiempo, ya que los desarrolladores de las criptomonedas apuestan a que llegue a ser el dinero virtual del futuro.

2.7.2 Viabilidad

La viabilidad de las criptomonedas no se basa en la generación de un ingreso, sino que depende más bien directamente de la participación de la comunidad (usuarios, mineros, y desarrolladores), consolidando su seguridad en la tecnología Blockchain, permitiendo la descentralización en cada transacción, a diferencia del dinero fiduciario que está ligado a un control centralizado, es decir, operan de forma independiente con respecto a bancos y gobiernos (La viabilidad de las criptomonedas)[10].

2.7.3 Versatilidad

La versatilidad que poseen las criptomonedas permite que cualquiera persona tenga opción de hacer uso de ellas. Todas las transacciones locales o internacionales son inmediatas y de cobros porcentuales mínimos, siendo el instrumento financiero que está convirtiendo a las criptomonedas en una ola de gran tamaño que comienza a generar expectativas.

2.8 Formas de Inversión en Criptomonedas

Los inversionistas de criptomonedas las consideran como una revolución del dinero, que puede convertirse en una gran oportunidad para elevar sus economías. Para considerarlas como medio de inversión se deben tener precisos y concisos los siguientes análisis.

2.8.1 Análisis Fundamental

Para todos los inversionistas de criptomonedas es importante el análisis fundamental para evaluar la viabilidad y el potencial de la misma de acuerdo con sus estados financieros. Se debe verificar si los números se ven bien, para poder estar seguros que la criptomoneda tiene buenos fundamentos y por lo tanto hacer opiniones propias si se quieren invertir en ella, y así asegurarse de estar libre de estafas. Existen ciertos pasos para que todo inversionista realice un buen análisis fundamental de la moneda que quiere invertir, como encontrar las fuentes adecuadas de información con el fin de evaluar una moneda. Se debe saber primero donde tener las principales fuentes de información, como el Libro blanco de la moneda, el Slack Channel o Blog de la moneda, Foros de la Comunidad, entre otros, y por último hacer preguntas en las comunidades sobre la moneda para determinar si se trata de una buena inversión (Una guía para el analisis fundamental para cryptocurrencies)[19].

2.8.2 Análisis Técnico

Para muchas personas las criptomonedas pueden valer mucho; para otros podría ser inútil. Sin embargo, el precio sigue moviéndose abruptamente y los inversionistas deben hacer análisis técnicos de cada pieza, dato o información ligada a las tendencias de los precios.

Para ello, se necesita el estudio de los gráficos, con intención de observar los precios y datos del volumen que se ha operado, y así, determinar las direcciones de las tendencias en el tiempo, buscando la predicción más cercana al cierto. El fundamento de este análisis es enfocarse al estudio de proyectar los movimientos de precios, y esperar el momento adecuado para invertir, por lo tanto, siempre que analicemos los movimientos de los precios en exclusiva se tratará de un análisis técnico del mercado (Qué es el Análisis técnico) [12].

2.9 Descripción de las Criptomonedas Seleccionadas

2.9.1 BITCOIN

- **Fecha de creación:** 3 de enero de 2009.
- **Algoritmo:** SHA-256.
- **Consenso:** Prueba de Trabajo (PoW).
- **Emisión máxima:** 21 millones de bitcoin cash (BTC).
- **Tiempo de bloque (Promedio):** 10 minutos.
- **Recompensa por bloque:** 12,5 BTC.
- **Ajuste de dificultad:** Cada 2016 bloques.
- **Página web:** <http://bitcoin.org>



Figura 2.4: Logo del Bitcoin

Fuente: Bitcoin.org

Bitcoin es la plataforma madre de la Blockchain, permitiendo el intercambio de usuario a usuario sin intermediarios gracias a la red P2P, son producidas por

las personas y empresas alrededor del mundo que dedican grandes recursos a la minería.

Fue creada en 2009 por el Seudónimo llamado Satoshi Nakamoto, presentada como la primera moneda descentralizada de la historia. Es utilizada para compra y venta de otras criptomonedas, que hoy en día sigue siendo la moneda de mayor valor y la más utilizada por los usuarios, es figurada como oro digital por la creciente elevación en su precio que ha tenido desde su aparición. Son minadas, y el creador de cada bloque es recompensado con 12.5 Bitcoin, pero este monto es reducido a la mitad cada, 210.000 bloques creados. Esta transición ocurrirá hasta llegar al monto limite creado por Nakamoto.

Con los años se han ido integrando nuevos desarrolladores a este proyecto de código abierto, contando ahora con más de 400 colaboradores que trabajan diariamente para enfrentar los obstáculos e incidentes que moldean a la moneda (Bitcoin)[2].

2.9.2 ETHEREUM

- **Fecha de creación:** 30 de julio de 2015.
- **Algoritmo:** Ethash.
- **Consenso:** Prueba de Trabajo (PoW) / Prueba de Participación (PoS) a futuro.
- **Emisión máxima:** Infinita (por ahora).
- **Tiempo de bloque (Promedio):** 15 segundos.
- **Recompensa por bloque:** 3 ethers (ETH).
- **Ajuste de dificultad:** 1 bloque.
- **Página web:** <https://www.ethereum.org>

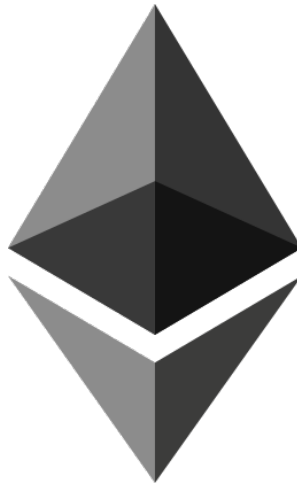


Figura 2.5: Ethereum

Fuente: Ethereum.org

Vitalik Buterin en 2013 comenzó el desarrollo de Ethereum pero su lanzamiento no llegó hasta el 2015; fue concebida por los usuarios como una versión mejorada de la criptomoneda Bitcoin, superando las limitaciones de su lenguaje de programación, convirtiéndose en la plataforma descentralizada más importante, proporcionando características avanzadas tales como, custodia sobre la Blockchain, límites de retiro, contratos financieros, mercado de juegos de azar y similares a través de un lenguaje de programación muy generalizado.

Ethereum está siendo utilizado por diversas compañías del mundo y en diversos sectores empresariales por la libertad que ofrecen sus contratos inteligentes de almacenar datos, enviar y recibir transacciones e incluso interactuar con otros contratos. Tiene el potencial para aumentar sustancialmente la eficiencia de la industria computacional, proporcionando a la red P2P a un impulso mayor, gracias a las considerables aplicaciones que no tiene vínculo con el dinero.

El minado anual de Ethereum es limitado a 18 millones y cuenta con un sistema de transacción coinbase, permitiendo que si dos mineros encuentran al mismo instante la solución de un bloque, los dos son recompensados, y se emite la moneda a la Blockchain de Ethereum, siendo la recompensa por bloque de 2 ETH (Qué es ethereum eth)[13].

2.9.3 LITECOIN

- **Fecha de creación:** 13 de octubre 2011.
- **Algoritmo:** Script.
- **Consenso:** Prueba de Trabajo (PoW).
- **Emisión máxima:** 84 millones de litecoins (LTC).
- **Tiempo de bloque (Promedio):** 2,5 minutos.
- **Recompensa por bloque:** 25 LTC.
- **Ajuste de dificultad:** 2016 bloques.
- **Página web:** <https://litecoin.com>



Figura 2.6: Litecoin

Fuente: Litecoin.com

Es la criptomoneda surgida como un complemento del Bitcoin y creada por un ex empleado de Google llamado Charlie Lee en 2011. A diferencia de muchos creadores de criptomonedas no conservó ningún Litecoin para sí mismo, sino que se decidió adquirirlo por sus propios medios como un usuario más. Litecoin es una criptomoneda que se desarrolló pensando en los pagos, generando bloques más rápidos y de comisiones muy bajas.

El algoritmo de su minado utiliza scrypt, y sus transacciones se realizan bajo su misma Blockchain que permite separar las firmas de confirmación de los datos de las transacciones, resultando mayor velocidad y seguridad de la misma. La ventaja de Litecoin es que su minado se puede hacer a través de una computadora convencional o alguna tarjeta de video que ayude a contar con el procesamiento adecuado, siendo su recompensa de 25 Litecoin por cada bloque confirmado (Litecoin: La plata de las criptomonedas)[11].

www.bdigital.ula.ve

C.C. Reconocimiento

Capítulo 3

Descripción de los Modelos de Predicción

3.1 Introducción

Todo inversionista tiene planes a futuro, siempre se encuentra en una constante incertidumbre de querer predecir los eventos para obtener resultados esperados. Pero no siempre sucede, porque pueden ocurrir una serie de eventos que desestabilicen el objetivo principal.

Para no experimentar con este tipo de situaciones es necesario pronosticar, evaluar y controlar cada evento, logrando disminuir tiempo o pérdidas de dinero, para ayudar a todo aquel inversionista que se encuentra en la incertidumbre de qué decisión tomar.

A continuación describiremos el funcionamiento de los modelos que serán ejecutados para predecir los valores futuros de las criptomonedas. El objetivo de este proyecto es presentar un conjunto de técnicas para los modelos de regresión lineal y series temporales, por lo tanto vamos a comenzar definiendo cada uno de los modelos predictivos, cuáles son sus características específicas y por qué es necesario desarrollar métodos específicos para su análisis.

3.2 Regresión Lineal

La regresión lineal tiene como objetivo modelar en forma matemática el comportamiento de una variable de respuesta en función de una o más variables independientes; puede ser representado por modelos estadísticos de regresión que permiten explicar y predecir su comportamiento, permitiéndonos analizar tanto contribuciones individuales como colectivas de un conjunto de variables independientes a la explicación de los cambios que se producen de la variable respuesta, la cual puede ser simple (si $k = 1$) o múltiple (si $k > 1$).

3.2.1 Regresión Lineal Simple

Sean dos variables X y Y , suponga que se quiere explicar el comportamiento de Y con base en los valores que toma X . Por consiguiente se mide el valor de Y sobre un conjunto de n valores de X , se obtienen n parejas de puntos $(x_1, y_1), (x_1, y_1), \dots, (x_n, y_n)$. A Y se le llama variable dependiente o la variable de respuesta, y a X se le conoce como variable independiente o variable regresora. Una manera de estudiar el comportamiento de Y con respecto a X mediante un modelo de regresión, es ajustando un modelo matemático de la siguiente forma:

$$Y = F(X)$$

Con ello, se puede ver si dado un valor de la variable independiente X es posible predecir el valor de Y .

A continuación observaremos varios tipos de relación en función de $F(X)$.

- **Lineal:** Cuando la función $F(X)$ es lineal.

$$F(X) = \beta_0 + \beta_1 X$$

Si $\beta_1 > 0$ hay una relación lineal positiva.

Si $\beta_1 < 0$ hay una relación lineal negativa.

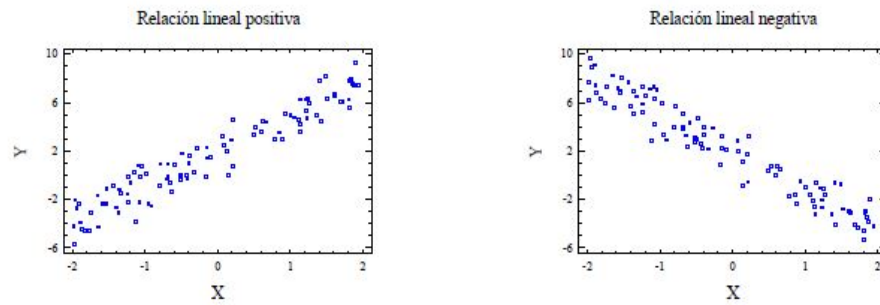


Figura 3.1: Gráfica de Relación Lineal

Fuente: slideserve.com

- **No Lineal:** Cuando la función $F(X)$ no lineal.

$$F(X) = \log(x), F(X) = x^2 + 3, \dots$$

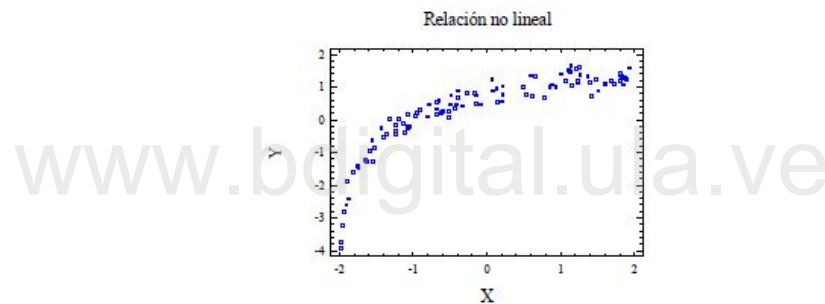


Figura 3.2: Gráfica de Relación No Lineal

Fuente: slideserve.com

- **Ausencia de Relación:** Cuando la función $F(X) = 0$.

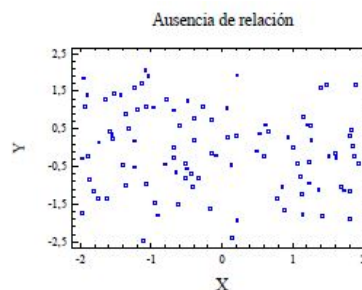


Figura 3.3: Gráfica de Ausencia de Relación

Fuente: slideserve.com

Si las variables X y Y , están relacionadas linealmente, donde cada valor de X la variable dependiente Y es una variable aleatoria. Es decir, que cada observación de Y puede ser descrita por la siguiente ecuación.

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Donde ε es un error aleatorio con media o valor esperado igual a cero $E(\varepsilon) = 0$. La varianza de ε , representado por σ^2 , es igual para todos los valores de X . β_0 y β_1 son parámetros desconocidos del modelo, donde β_0 es el punto en el cual la línea recta intercepta o cruza el eje Y , y β_1 es la pendiente de la línea. Por lo tanto, para tener bien especificado la ecuación que relaciona las dos variables será necesario estimar β_0 y β_1 mediante el método de mínimos cuadrados.

3.2.2 Criterio de los Mínimos Cuadrados

Este método emplea los datos de la muestra para estimar los parámetros del modelo de regresión que minimiza los errores de ajuste. En otras palabras, debemos encontrar la recta que pasa más cerca de todos los puntos. Partiendo de la ecuación lineal más sencilla:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

El término del error ε_i puede ser entendido, como la distancia que existe entre el valor observado Y_i , y el correspondiente valor estimado $\hat{Y}$. El objetivo de la estimación por Mínimos Cuadrados Ordinarios es, precisamente, minimizar el sumario de todas esas distancias al cuadrado; es decir:

$$S = \sum_{i=1}^n (\varepsilon_i)^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2$$

Derivando esta expresión respecto a los coeficientes $\hat{\beta}_0$ y $\hat{\beta}_1$ e igualando a cero obtenemos el sistema de ecuaciones normales:

$$\sum_{i=1}^n Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_i \Rightarrow \bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$$

$$\sum_{i=1}^n Y_i X_i = \hat{\beta}_0 \sum_{i=1}^n X_i + \hat{\beta}_1 \sum_{i=1}^n X_i^2$$

Donde n representa el tamaño muestral y $\bar{X}$ e $\bar{Y}$ representan las medias de dicha variables. Resolviendo el sistemas de ecuaciones obtenemos la solución para los parámetros $\hat{\beta}_0$ y $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

3.2.3 Pruebas de Hipótesis en la Regresión

En cualquier análisis de regresión no basta hacer los cálculos, sino que es necesario evaluar qué tanto el modelo (la línea recta) explica la relación entre X y Y . Una primera forma de hacer esto es probar una serie de hipótesis.

Por lo general, la hipótesis de mayor interés plantea que la pendiente es significativamente diferente de cero. Esto se logra al probar las siguientes hipótesis.

H_0 : Hipótesis Nula.

H_A : Hipótesis Alternativa.

Al aplicar uno de estos criterios, se rechaza H_0 si el valor absoluto de este estadístico es mayor que el correspondiente valor crítico obtenido de tablas, es decir, se rechaza H_0 si:

$$F = \left(\frac{\alpha}{2}, n - 2 \right)$$

También se utiliza como criterio de rechazo la comparación de la significancia observada (p-valor o valor-p) contra la significancia predefinida ($\alpha = 0,05$), entonces se rechaza H_0 si $\text{valor} - p < \alpha$.

3.2.4 Análisis de Varianza

La idea general de esta técnica es separar la variación total en las partes que contribuye cada fuente de variación del modelo.

Una de las propiedades importantes del método de mínimos cuadrados, es que si igualamos cada una de las diferencias al cuadrado y la sumamos obtendríamos una relación existente entre las tres sumas. Para analizar la significancia del modelo.

$$SCT = SCR + SCE$$

Suma de Cuadrado de la Regresión (SCR):

Cuando los valores de $\hat{Y}_i$ medidos en la línea de regresión de los valores de $\bar{Y}$ obtenemos la suma de cuadrados debido a la regresión, con 1 grado de libertad, puesto que no basta conocer los valores de β_0 y β_1 para poder calcular el valor de la suma.

Suma de Cuadrado Residual (SCE):

Es una medida del error incurrido al usar la ecuación de regresión para calcular los valores de la variable dependiente en la muestra, consta de $n - 2$ grados de libertad. Debemos restar al número total de elementos dos grados de libertad correspondientes a los coeficientes de β_0 y β_1 .

Suma de Cuadrado Residual (SCE):

Es un valor importante al usar $y?$ para estimar los valores del modelo, donde su número de grado de libertad es de $n - 1$. El último sumando es combinación lineal de los anteriores, por definición de la media, la suma de las diferencias es igual a cero.

Las sumas de cuadrados divididas entre sus respectivos grados de libertad se llaman cuadrados medios. Los dos de mayor importancia son los cuadrados medio de la regresión (CM_R) y el cuadrado medio residual (CM_E), permitiendo así construir una tabla de análisis de varianza para regresión lineal simple, ilustrando mejor los resultados.

Variación	Suma de Cuadrados	Grados de Libertad	Cuadrado Medio	F
Regresión	$SCR : \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2$	1	$CM_R = \frac{SCR}{1}$	$F = \frac{CM_E}{CM_R}$
Residual	$SCE : \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$	$n - 2$	$CM_E = \frac{SCE}{n-2}$	
Total	$SCT : \sum_{i=1}^n (Y_i - \bar{Y})^2$	$n - 1$		

Tabla 3.1: Tabla de Análisis de Varianza de Regresión Lineal Simple

3.2.5 Calidad del Ajuste

Para valorar la calidad de ajuste es preciso recordar que el objetivo de la regresión es explicar los cambios que se dan en la variable dependiente; cuando mayor sea el porcentaje de cambio mayor será el ajuste obtenido, para esto se introducirá un nuevo índice estadístico conocido como Coeficiente de Determinación R^2 .

Coeficiente de Determinación R^2 :

Se interpreta como la proporción de la variabilidad en los datos de Y que es explicada por el modelo y ronda entre el intervalo de $0 < R^2 \leq 1$, cuantificando el porcentaje de variabilidad presente en los datos explicado por el modelo. Por ello, son deseables valores próximos a 1, permitiendo que la calidad del modelo sea satisfactorio.

$$R^2 = \frac{SCR}{SCT}$$

Coeficiente de Determinación R_{aj}^2 :

Cuando hay muchos términos en el modelo, el estadístico de R_{aj}^2 se prefiere en lugar de R^2 , puesto que este último es engañoso al incrementarse en forma artificial con cada termino que se agrega al modelo; en cambio el R_{aj}^2 incluso baja de valor cuando el término que se agrega no aporta en nada. De esta forma, para interpretar estos coeficientes se cumple que $0 < R_{aj}^2 \leq R^2 \leq 1$. En general, para fines de predicción se recomienda un coeficiente de determinación ajustado de al menos 0.7.

$$R_{aj}^2 = \frac{CM_{Total} - CM_E}{CM_{Total}} = \frac{(n-1)R^2 - k}{n-k-1}$$

Coefficiente de Correlación:

Es bien conocido que el coeficiente de correlación, r , mide la intensidad de la relación lineal entre X y Y . Si r es próximo a -1 , tendremos una relación lineal negativa fuerte, si r es próximo a cero diremos que no hay correlación lineal, y finalmente si r es próximo a 1 tendremos una relación lineal positiva fuerte.

Error estándar de Estimación:

Es una medición sobre la calidad del ajuste de un modelo, representado como el error estándar de estimación; que a medida que el modelo ajuste mejor la suma de cuadrados del error será menor y en consecuencia el error estándar de estimación también será menor.

$$\hat{\sigma} = \sqrt{\frac{SCR}{n-2}}$$

Media del Error Absoluto:

Otra forma de medir la calidad del ajuste es a través de la media del valor absoluto de los residuos, mientras mejor sea el ajuste los residuos serán más pequeñas y en consecuencia también la media del error absoluto tenderá a ser más pequeña. Se puede ver como una medición para observar cuando falla en promedio el modelo, al hacer la estimación de la variable de respuesta.

$$mea = \frac{\sum_{i=1}^n |\varepsilon_i|}{n}$$

3.2.6 Análisis de Residuales

Un análisis adecuado de los residuos proporciona información adicional sobre la calidad del ajuste del modelo de regresión, permitiendo verificar si el modelo es adecuado. Las gráficas que suelen hacerse para completar el diagnóstico del modelo consiste en:

Gráfica de Residuales en Función de la Variable Independiente:

Se define como la diferencia entre la respuesta observada y la respuesta predicha por el modelo, lo cual permite hacer un diagnóstico más directo de su calidad. La gráfica debe mostrar un patrón similar a una franja horizontal de puntos dispersamente distribuidos, para lograr cumplir una representación adecuada de relación entre las variables.

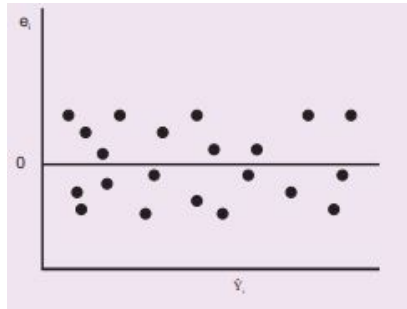


Figura 3.4: Gráfica de los Residuos en Función de la Variable Independiente

Fuente: researchgate.net

Grafica de Residuales Estandarizados:

Los residuales estandarizados se obtienen al dividir los residuos entre su respectiva desviación estándar. Si el modelo es correcto los residuales deberían aparecer en cualquier grafico como una variación aleatoria alrededor de cero como el grafico anterior de residuos. Se usa para observar la existencia de valores atípicos o influyentes, como se ilustra en la figura 3.5.

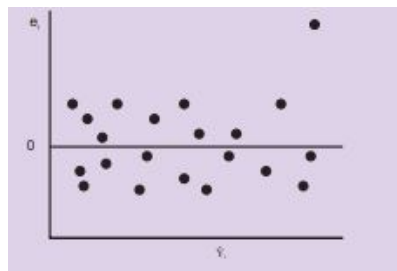


Figura 3.5: Gráfica de los Residuales Estandarizados

Fuente: researchgate.net

Gráfica de Probabilidad Normal:

Consiste en graficar los residuos en papel de probabilidad normal que se incluye en todos los paquetes estadísticos. Esta grafica del tipo $X - Y$ tiene las escalas de tal manera que si los residuos siguen una distribución normal al graficarlos, tienden a quedar alineados en una línea recta como se ilustra en la figura 3.6.

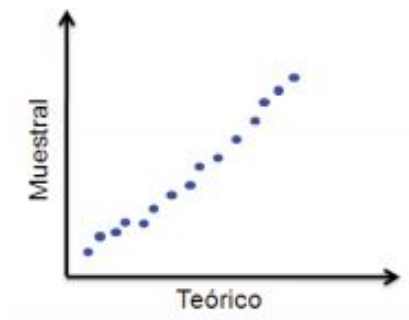


Figura 3.6: Gráfica de Probabilidad Normal

Fuente: es.slideshare.net

3.2.7 Verificación del Supuesto de Independencia

Para verificar la suposición del supuesto de independencia se suelen graficar los residuos contra la orden en el que se obtuvieron los datos. La suposición se cumple si los puntos de la gráfica no siguen ningún patrón bien definido.

La violación de este supuesto se le conoce como autocorrelación. Si esto sucede, las pruebas estadísticas del modelo de regresión no serán válidas, ya no tendrán la varianza mínima. La manera de validar el cumplimiento del supuesto de independencia es con la prueba de Durbin-Whatson.

Prueba de Durbin-Watson:

Permite diagnosticar la presencia de correlación entre los residuos que es una posible manifestación de la falta de independencia.

Sea ρ el parámetro que representa la correlación entre residuos consecutivos ($\text{corr}(\varepsilon_i, \varepsilon_{i-1}) = \rho; i = 2, 3, \dots, n$). La hipótesis en la prueba de Durbin-Watson es:

$h_0: \rho = 0$: Evidencia estadística que hay independencia de los términos del error.

$h_A: \rho \neq 0$: Evidencia estadística que los términos de error están autocorrelacionados.

El estadístico de Durbin-Watson es:

$$d = \frac{\sum_{i=2}^n (\varepsilon_i - \varepsilon_{i-1})^2}{\sum_{i=1}^n (\varepsilon_i)^2}$$

Donde los $\varepsilon_i, i = 1, 2, \dots, n$, son los residuos ordenados en el tiempo. La decisión de la hipótesis planteada consiste en:

Si $d < d_L$: Se rechaza H_0

Si $d > d_U$: No se rechaza H_0

Si $d_L < d < d_U$: Sin decisión

Donde d_L y d_U son valores expresados en tablas.

Prueba de Falta de Ajuste:

Con las pruebas que hemos visto hasta aquí suelen ser suficientes para evaluar la calidad de ajuste del modelo, aunque es posible desarrollar una prueba adicional conocida como prueba de falta de ajuste; consta de evaluar si una relación curvilínea podría ajustarse mejor a los datos que un modelo lineal. La idea de esta prueba es descomponer la SCE en dos partes:

$$SCE = SCE_R + SCE_F$$

SCE_R : Suma de cuadrados atribuibles a un error.

SCE_F : Suma de cuadrados atribuibles a la falta de ajuste del modelo.

Para estimar SCE_F será necesario que para al menos un valor X_i haya varias observaciones de Y . Estos dos componentes son utilizados para construir un estadístico

de prueba F particular con el fin de evaluar la siguiente hipótesis:

H_0 : La relación es lineal.

H_A : La relación no es lineal.

$$F_o = \frac{SCE_F/(m-2)}{SCE_R/(n-m)} = \frac{CM_F}{CM_R}$$

Este estadístico tiene una distribución F con $(m-2)$ y $(n-m)$ grados de libertad en el numerador y denominador. Se rechaza H_0 si $F_o > F_{(\sigma, m-2, n-m)}$.

El resultado de esta hipótesis debe ser comparado con el resto de los criterios sobre la calidad de los modelos vistos antes. Si están orientados en el mismo sentido, entonces la prueba de falta de ajuste será un elemento adicional.

3.2.8 Predicción a Partir del Modelo

Una de las aplicaciones más útiles de un modelo de regresión es predecir o pronosticar nuevas o futuras observaciones de Y . Una estimación puntual de la observación futura de y_0 en el punto x_0 está dada por:

$$\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

Para predecir por intervalos, como la nueva observación es independiente de las observaciones utilizadas para ajustar el modelo de regresión. Sea y_0 la observación futura en x_0 y sea $\hat{y}_0$ el estimador puntual para y_0 ; de ellas se define el error de esta estimación como:

$$\hat{\varepsilon}_0 = y_0 - \hat{y}_0$$

Esta variable, $\hat{\varepsilon}_0$, se distribuye de manera normal con media cero y varianza:

$$\begin{aligned} V(\hat{\varepsilon}_0) &= V(y_0 - \hat{y}_0) = V(y_0) + V(\hat{y}_0) \\ &= \sigma^2 + \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{X})^2}{\sum_{i=1}^n (x_i - \bar{X})^2} \right] \\ &= \sigma^2 \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{X})^2}{\sum_{i=1}^n (x_i - \bar{X})^2} \right] \end{aligned}$$

Dada esta varianza para el error de predicción, y estimando σ^2 con el CM_E , entonces el intervalo de predicción al $(1 - \alpha) * 100$ con respecto a la observación futura en x_0 está dada por:

$$\left[\hat{y}_0 \pm t_{(\alpha/2, n-2)} \sqrt{CM_E \left[1 + \frac{1}{n} + \frac{(x_0 - \bar{X})^2}{\sum_{i=1}^n (x_i - \bar{X})^2} \right]} \right]$$

3.3 Series de Tiempo

Una serie de tiempo es un conjunto de observaciones producidas en determinados momentos, ordenados cronológicamente durante un periodo, semanal, mensual, trimestral o anual, generalmente a intervalos iguales. Para describir una serie de tiempo se debe identificar el comportamiento de los datos, que pueden observarse gráficamente para ser identificadas, pero no en todos los casos es posible distinguir las particularidades que cada una puede contener. En estudios antecesores la experiencia basada en muchos ejemplos de series de tiempo ha revelado que existen ciertos movimientos o variaciones, características que pueden medirse y observarse por separado. Estas variaciones son definidas por cuatros tiempos principales, llamados a menudo componente de una serie de tiempo:

3.3.1 Componentes de una Serie Temporal

Se dice que una serie de tiempo puede descomponerse en cuatro componentes cinco si se considera una constante llamada nivel que no son directamente observables, de los cuales únicamente se puede hacer estimaciones. Estos cuatros componentes son:

- **Tendencia:** Representa el comportamiento predominante de la serie, resultado de factores que afectan a los datos, tienden aumentar o disminuir en un periodo muy largo.
- **Cíclicas:** Son movimientos ascendentes y descendentes recurrentes respecto a la tendencia con una duración de varios años.

- **Estacionales:** Son movimientos ascendentes y descendentes respecto a la tendencia que se consuman dentro de un año y se repiten anualmente.
- **Aleatorio:** Son movimientos erráticos que no siguen un patrón específico son imprescindibles. Este comportamiento representa todos los tipos de movimiento de una serie de tiempo que no son tendencia, variaciones estacionales ni fluctuaciones cíclicas.

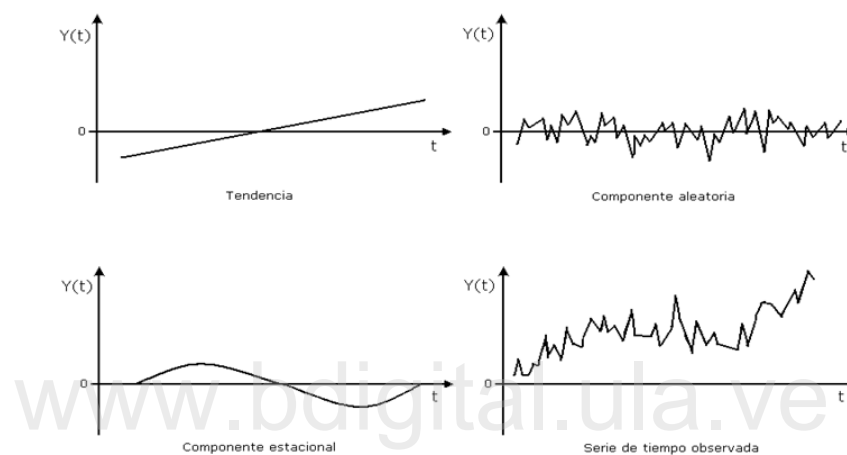


Figura 3.7: Componentes de una Serie Temporal

Fuente: Monografias.com

3.3.2 Clasificación Descriptiva de las Series Temporales

Las series temporales se pueden clasificar en:

- **Estacionarias:** Una serie es estacionaria cuando es estable, es decir, cuando la media y varianza no cambia a lo largo del tiempo y la covarianza solo es función del rezago.
- **No Estacionarias:** Son series cuyas características de media, varianza y covarianza cambian en el tiempo. Los cambios en la media determinan una tendencia a crecer o decrecer a largo plazo, por lo que la serie no oscila alrededor de un valor constante.

3.3.3 Procesos Estocásticos

En el análisis de series temporales el objetivo es utilizar la teoría de procesos estocásticos para determinar qué proceso estocástico ha sido capaz de generar la serie temporal bajo estudio con el fin de caracterizar el comportamiento de la serie y predecir en el futuro.

Procesos Estocástico Estrictamente Estacionario

Un proceso estocástico es estrictamente estacionario si su media, varianza y covarianza permanecen iguales a lo largo del tiempo, manteniendo sus características estadísticas iguales sin importar el momento que se midan.

$$\text{Media: } E(Y_t) = E(Y_{t+k}) = \mu$$

$$\text{Varianza: } Var(Y_t) = V(Y_{t+k})^2 = \sigma^2$$

$$\text{Covarianza: } E[(Y_t - \mu)(Y_{t+k} - \mu)] = E[(Y_{t+m} - \mu)(Y_{t+m+k} - \mu)] = \gamma_k$$

Proceso Estocástico Débilmente Estacionarios

Un proceso estocástico se dice que es débilmente estacionario si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos depende solamente de la distancia o rezago entre estos dos periodos de tiempo y no del tiempo en el cual se ha calculado su covarianza.

$$\text{Media: } E(Y_t) = \mu$$

$$\text{Varianza: } Var(Y_t) = E(Y_t - \mu)^2 = \sigma^2$$

$$\text{Covarianza: } E[(Y_t - \mu)(Y_{t+k} - \mu)] = \gamma_k$$

Proceso Puramente Aleatorio

Es un proceso puramente aleatorio, si las variables conforman una secuencia independientes e idénticamente distribuidas, con media y varianza constante y covarianza cero.

$$\text{Media: } E(Y_t) = \mu$$

$$\text{Varianza: } Var(Y_t) = E(Y_t - \mu)^2 = \sigma^2$$

$$\text{Covarianza: } E[(Y_t - \mu)(Y_{t+k} - \mu)] = 0$$

3.3.4 Ruido Blanco (White Noise)

Es una señal aleatoria que se caracteriza por el hecho de que sus valores de señal en dos tiempos diferentes no guardan correlación estadística. Además, se trata de un proceso en el que todas sus variables son independientes con media cero, varianza constante y covarianza nula.

$$\text{Media: } E(Y_t) = 0$$

$$\text{Varianza: } Var(Y_t) = V(Y_t - \mu)^2 = \sigma^2$$

$$\text{Covarianza: } E[(Y_t - \mu)(Y_{t+k} - \mu)] = 0$$

3.3.5 Camino Aleatorio (Random Walks)

Se dice que el proceso Y_t es un camino al azar si:

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

Donde ε_t es la variable aleatoria de probabilidad para tomar el siguiente paso y t es el intervalo de tiempo entre pasos subsecuentes, a medida que se da un paso solo dependerá de Y_t y no de algunas posiciones previa. Por ejemplo:

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

$$Y_{t-1} = Y_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

$$Y_{t-2} = Y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

Así sucesivamente, expresado de la siguiente forma:

$$Y_t = Y_0 + \sum \varepsilon_t$$

Este proceso puede o no incluir un parámetro de variación δ (DRIFT), en torno al cual gira $\varepsilon_t = Y_t - Y_{t-1}$. Si se toma diferencia de primer orden puede transformarse en un proceso puramente aleatorio.

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

En general sus propiedades son:

$$E(Y_t) = E\left(Y_0 + \sum \varepsilon_t\right) = Y_0 + T\delta$$

$$Var(Y_t) = V\left(Y_0 + \sum \varepsilon_t\right) = Y_0 + T\sigma^2$$

Por lo tanto, el proceso no es estacionario su media y varianza se incrementa en el tiempo.

3.3.6 Función de Autocovarianza (ACV)

Es la covarianza entre cualquier par de componentes de Y_t separados entre sí por un intervalo temporal o retardo h .

Sus propiedades son:

$$\gamma_0 = Var(Y_t) > 0$$

$$|\gamma_h| \leq \gamma_0$$

$$\gamma_h = \gamma_{-h} \quad (simetria)$$

Γ_h Matriz semidefinida positiva

Aunque γ_h puede depender de h , pero no depende de los momentos concretos a los que se refieran los componentes de Y_t considerados.

$$\gamma_h = Cov(Y_t, Y_{t-h}) = E[(Y_t - \mu)(Y_{t+k} - \mu)]$$

3.3.7 Función de Autocorrelación

Define la relación mutua existente entre valores de una serie de tiempo en diferentes periodos, permitiéndonos analizar los patrones de tendencia, estacionalidad e irregularidades que puedan presentar las series. Existen dos formas de estudiar la autocorrelación que sirven para conocer tanto la existencia de correlación entre los datos de la serie como el grado o intensidad y el tipo de la misma.

Función de Autocorrelación Simple (FAS o PAC)

Mide la relación lineal entre las observaciones de una serie de datos Y_t , distanciados en un lapso de tiempo h , que se le conoce como retardo o retraso. La autocorrelación simple de un proceso Y_t estacionario se representa con el símbolo ρ_h y se define como:

$$\rho_h = \frac{Cov(Y_t, Y_{t-h})}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t+h})}} = \frac{\gamma_h}{\gamma_0}$$

Sus propiedades son:

$$\rho_0 = 1$$

$$|\rho_h| \leq 1$$

$$\rho_h = \rho_{-h} \quad (\text{simetria})$$

$$S_\rho^2 = \frac{1}{n} \left(1 + 2 \sum \rho_h^2 \right)$$

El cálculo del coeficiente de autocorrelación simple r_h entre observaciones que se encuentran a h periodos de distancia se realiza por medio de la siguiente expresión:

$$r_h = \frac{\sum_{t-s}^T (Y_t - \bar{Y})(Y_{t-s} - \bar{Y})}{\sum_{t-s}^T (Y_t - \bar{Y})^2}$$

Donde:

r_h : Coeficiente de autocorrelación simple para un retraso de h periodos.

$\bar{Y}$: Media de los valores de la serie.

Y_t : Observación en el periodo t .

Y_{t-s} : Observación en el periodo con h retrasos.

N : número total de periodos.

Función de Autocorrelación Parcial (FAP o AC)

Se puede describir la autocorrelación parcial como la estimación de la autocorrelación simple para el mismo retardo h . La diferencia entre estos dos tipos de autocorrelación se halla en que la autocorrelación simple brinda para un retardo h , mientras que la autocorrelación parcial brinda solo la relación para la diferencia estricta en h retrasos de tiempo.

$$\phi_h = \frac{cor(Y_t, Y_{t+h})}{(Y_{t+1}, Y_{t+2}, \dots, Y_{t+h-1})}$$

Mide la contribución de la observación Y_t en la explicación de Y_{t+h} .

El grado de relación, tanto para la autocorrelación simple y parcial, se mide mediante el coeficiente de autocorrelación, que también se distingue entre simple y parcial. Ambos coeficientes de autocorrelación varían entre $+1$ y -1 . Además permite determinar el orden de los procesos autoregresivos.

El correlograma o función de autocorrelación es una gráfica de las autocorrelaciones para varios retrasos de tiempo de una serie de tiempo.

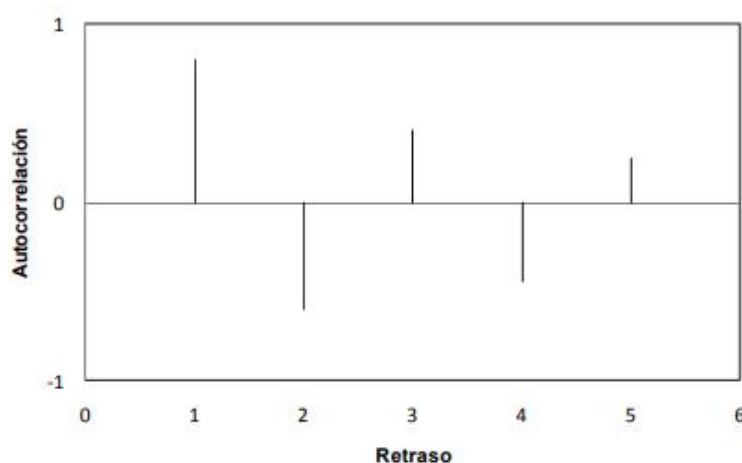


Figura 3.8: Correlograma de Serie de Tiempo

Fuente: scielo.org.ar

El correlograma nos muestra en la escala horizontal de la parte de abajo cada uno de los retrasos de tiempo que son de interés, mientras que la escala vertical del lado izquierdo muestra el posible rango de un coeficiente de autocorrelación de -1 y $+1$. La línea horizontal a la mitad de la gráfica representa autocorrelaciones de cero.

Se calcula como el último coeficiente ϕ_{hh} de la regresión múltiple de Y_t respecto de $(Y_{t+1}, Y_{t+2}, \dots, Y_{t+h-1})$:

$$Y_t = \phi_{21}Y_{t-1} + \phi_{22}Y_{t-2} + \varepsilon_t$$

$$Y_t = \phi_{31}Y_{t-1} + \phi_{32}Y_{t-2} + \phi_{33}Y_{t-3} + \varepsilon_t$$

Y así sucesivamente o a través de las ecuaciones de Yule-Walker

$$Y_t = \phi_{h1}Y_{t-1} + \phi_{h2}Y_{t-2} + \phi_{h3}Y_{t-3} + \dots + \phi_{hh}Y_{t-h} + \varepsilon_t$$

Significancia Estadística Conjunta de los Coeficientes de Autocorrelación

Se establece a través de la siguiente prueba de hipótesis:

$H_0 : \rho_1 = \rho_2 = \dots = \rho_p = 0$ (Los coeficientes de autocorrelación son nulos, NO HAY AUTOCORRELACION SERIAL).

H_A : Al menos $\rho_j \neq 0$ (HAY AUTOCORRELACIÓN SERIAL).

La determinación de esta hipótesis se puede realizar a través de los siguientes estadísticos:

- **Estadístico Box-Pierce:** Utiliza el estadístico Q_h para probar la hipótesis nula de que las primeras autocorrelaciones de una serie tiempo débilmente estacionaria son cero.

$$Q_h = n \sum_{h=1}^m \hat{\rho}_h^2 \sim X_m^2$$

- **Estadístico Ljung-Box:** Esta prueba permite probar en forma conjunta de que todos los coeficientes de autocorrelación son simultáneamente iguales a cero, en otras palabras son independientes y está definida como:

$$LB = n(n+2) \sum_{h=1}^m \left(\frac{\hat{\rho}_h^2}{n-h} \right) X_m^2$$

Donde n es el tamaño de la muestra y m longitud del rezago.

3.3.8 Proceso Lineal Estacionario

Modelos Autorregresivos AR(p)

Los modelos autorregresivos se basan en la idea que el valor actual de la serie Y_t , puede explicarse en función de p valores pasados ($Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$), donde p determina el número de rezagos necesarios para pronosticar en valores futuro de la misma. El modelo autorregresivo de orden p está dado por:

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

Expresado en términos del operador de retardos:

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

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

Donde ε_t es un proceso de ruido blanco con media cero y varianza constante, $\phi_p(L)$ es el polinomio autorregresivo.

- **Proceso Autoregresivo de Orden Uno AR (1):** Utiliza el estadístico Q_h para probar la hipótesis nula de que las primeras autocorrelaciones de una serie tiempo débilmente estacionaria son cero.

En los procesos AR(1) la variable Y_t está determinado únicamente por un valor pasado (Y_{t-1}).

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

Se puede expresar usando la notación del operador de retardo:

$$(1 - \phi L)Y_t = \delta + \varepsilon_t$$

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

Donde δ es una constante que representa la variación no explicada por el modelo, ϕ_1 es el polinomio de orden p del componente autorregresivo y ε_t es un término del error estocástico.

Para cualquier valor del parámetro, el proceso AR(1) es estacionario si cumple las siguientes propiedades:

$$E(Y_t) = E(\delta + \phi_1 Y_{t-1} + \varepsilon_t) = \mu = \frac{\delta}{1 - \phi_1} \quad \text{para } \delta \neq 0 \quad \text{y} \quad |\phi_1|$$

$$Var(Y_t) = \phi_1^2 Var(Y_{t-1}) + Var(\varepsilon_t) = \gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2}$$

$$Cov(Y_t, Y_{t-h}) = \gamma_h = \phi_1^h \gamma_0 = \phi_1^h \gamma_{h-1}$$

$$Cor(Y_t, Y_{t-h}) = \rho_h = \frac{\gamma_h}{\gamma_0} = \phi_1^h = \phi_1 \rho_{h-1}$$

Haciendo uso de sustituciones repetidas, puede verse que:

$$Y_t = (\delta + \delta\phi_1 + \delta\phi_1^2 + \cdots) + (\varepsilon_t + \phi_1\varepsilon_{t-1} + \phi_1^2\varepsilon_{t-2} + \cdots) = \frac{\delta}{1 - \phi_1} + \sum_{s=0}^{\infty} \phi_1^s \varepsilon_{t-s}$$

Recordemos que suponemos que el proceso es no anticipante, es decir, el futuro no incluye en el pasado. Por lo que un modelo AR(1) es una versión restringida de un modelo general de medias móviles.

Las características más relevantes del modelo AR(1) son:

- El modelo AR(1) es siempre invertible.
- El modelo AR(1) es estacionario siempre que se cumpla $|\phi| < 1$.

- La función de autocorrelación total no se anula, tendrá un comportamiento amortiguado hacia cero con todos los valores positivos, en caso de que $\phi > 0$, o bien alternando el signo, comenzando con negativos, si $\phi < 0$.
 - La función de autocorrelación parcial se anula para retardos superiores a uno (el orden del modelo).
- **Proceso Autoregresivo de Orden Dos AR(2):** En los procesos AR(2) la variable Y_t está determinado por el valor pasado y el anterior a este.

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

Igual que el proceso de AR(1), pero este es determinado por dos operadores de retardos.

$$(1 - \phi L - \phi_2 L^2)Y_t = \delta + \varepsilon_t$$

Para que un proceso AR(2) sea estacionario tiene que cumplir con las siguientes propiedades:

$$|\phi_2|, \quad (\phi_1 + \phi_2) < 1, \quad (\phi_1 - \phi_2) < 1$$

$$E(Y_t) = \mu = \frac{\delta}{1 - \phi_1 - \phi_2} \quad \text{para } \delta \neq 0$$

$$Var(Y_t) = \gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1^2 - \phi_2^2}$$

$$Cov(Y_t, Y_{t-h}) = \gamma_h = \phi_1 \gamma_{h-1} + \phi_2 \gamma_{h-2}$$

$$Cor(Y_t, Y_{t-h}) = \rho_h = \frac{\gamma_h}{\gamma_0} = \phi_1 \rho_{h-1} + \phi_2 \rho_{h-2}$$

Un proceso AR(2) se puede identificar mediante el correlograma, ya que se detienen abruptamente en un punto. La grafica tendrá tantas barras como orden tenga el proceso. Son los picos significativos en los rezagos p , que en este caso sería dos.

Modelos de Media Móviles MA(q)

El proceso AR recién expuesto no es el único mecanismo que puede generar a Y_t . También puede ser expresado por los modelos de medias móviles de orden finito q . Siendo una aproximación natural al modelo lineal general antes expuesto en los procesos autoregresivos.

Enfocaremos nuestra atención al modelo lineal general en términos de operadores de retardo para hallar la ecuación que ejecutará los modelos de media móviles.

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

$$Y_t = \frac{1}{(1 - \theta_q L)} \varepsilon_t$$

El coeficiente $\frac{1}{(1 - \theta_q L)}$ podemos expresarlo como un polinomio finito:

$$\frac{1}{(1 - \theta_q L)} = 1 - \theta L - \theta_2 L^2 - \dots$$

Sustituyendo obtendremos un modelo AR de orden infinito:

$$Y_t = \frac{1}{(1 - \theta_q L)} \varepsilon_t = \delta + (1 - \theta L - \theta_2 L^2 - \dots) \varepsilon_t$$

Finalmente la ecuación que ejecutará los modelos de media móviles es la siguiente:

$$Y_t = \delta + \theta \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

Donde δ es una constante que representa la variación no explicada por el modelo, θ_q es el polinomio de orden q del componente de medias móviles y ε_t es un término de error (ruido blanco).

- **Proceso de Media Móviles de Orden Uno MA(1):** Determina el valor de Y en el momento t en función de la innovación contemporánea y su primer retardo:

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

Este modelo también se puede escribir en función del operador de retardo:

$$Y_t = \delta + (1 - \theta_1 L)\varepsilon_t$$

Es estacionario si cumple las siguientes propiedades:

$$E(Y_t) = E(\delta + \varepsilon_t - \theta_1 \varepsilon_{t-1}) = \mu_t = \delta \quad \text{para } \delta \neq 0 \quad \text{y} \quad |\theta_1| < 1$$

$$Var(Y_t) = \gamma_0 = (1 - \theta_1^2)\sigma_\varepsilon^2$$

$$Cov(Y_t, Y_{t-h}) = \gamma_h = \theta_1 \sigma_\varepsilon^2 \quad \text{si } h = 1; \quad \text{si } h \geq 2; \quad \gamma_h = 0$$

$$Cor(Y_t, Y_{t-h}) = \rho_h = \frac{\gamma_h}{\gamma_0} = \frac{\theta_1}{(1 + \theta_1^2)} \quad \text{para } h = 1; \quad \text{si } h \geq 2; \quad \rho_h = 0$$

Las características más relevantes del modelo MA(1) son:

- Siempre es estacionario.
- Para ser invertible, es necesario que $|\theta_1| < 1$.
- El modelo olvida la correlación con periodos distintos al inmediatamente anterior y el correlograma solo tendrá un punto significativo.
- La función de autocorrelacion parcial no se anula, pero tendrá un comportamiento amortiguado hacia cero.

- **Proceso de Media Móviles de Orden Uno MA(2):** Para analizar este proceso se tendrá en cuenta la siguiente ecuación:

$$Y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

Otra forma alterna de escribir este proceso es a través de operadores de retardo:

$$Y_t = \delta + (1 - \theta_1 L - \theta_2 L^2)\varepsilon_t$$

Para que un proceso MA(2) sea estacionario tiene que cumplir con las siguientes propiedades:

$$E(Y_t) = E(\delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}) = \mu_t = \delta \quad \text{para } \delta \neq 0; \quad |\theta_1| < 1 \quad \text{y} \quad |\theta_2| < 1$$

$$Var(Y_t) = \gamma_0 = (1 - \theta_1^2 - \theta_2^2) \sigma_\varepsilon^2$$

$$Cov(Y_t, Y_{t-h}) = \gamma_h = \sigma_\varepsilon^2(-\theta_1 + \theta_1 \theta_2) \quad \text{si } h = 1; \quad \text{y} \quad h = 2; \quad \gamma_h = \theta_2 \sigma_\varepsilon^2;$$

$$\text{si } h > 2; \quad \gamma_h = 0$$

$$Cor(Y_t, Y_{t-h}) = \rho_h = \frac{\gamma_h}{\gamma_0} = \frac{\theta_1(1 + \theta_2)}{(1 + \theta_1^2 \theta_2^2)} \quad \text{para } h = 1; \quad \rho_h = \frac{-\theta_2}{(1 + \theta_1^2 \theta_2^2)}$$

$$\text{si } h > 2; \quad \rho_h = 0$$

Procesos Autorregresivos de Medias Móviles ARMA(p,q)

Los modelos ARMA son procesos estacionarios los cuales están basados en la combinación de los procesos autorregresivos AR(p) y los procesos de medias móviles MA(q). La variable Y_t queda como función de sus valores pasados y sus errores estimados cumpliendo con las condiciones antes descritas.

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

Donde δ es un término constante, ϕ_p son los parámetros autorregresivos, θ_q son los parámetros de medias móviles y ε_t es un proceso de ruido blanco.

El modelo en términos del operador de retardos queda de la siguiente manera:

$$(1 - \phi_1 L - \cdots - \phi_p L^p) Y_t = (1 - \theta_1 L - \cdots - \theta_q L^q) \varepsilon_t$$

$$\phi_p(L) Y_t = \theta_q(L) \varepsilon_t$$

Siendo $\phi_p(L)$ el polinomio autorregresivo y $\theta_q(L)$ es el polinomio de media móviles.

El modelo ARMA(p,q) va a compartir las características de los modelos AR(p) y MA(q) ya que contienen ambas estructuras a la vez. El modelo ARMA(p,q) tiene media cero, varianza constante y finita con una función de autocorrelacion infinita que decrece rápidamente hacia cero.

• **Procesos Autorregresivos de Medias Móviles de Orden ARMA(1,1):**

El modelo ARMA(1,1) tiene la siguiente estructura:

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

Donde Y_t está en función de su pasado hasta el primer retardo, la innovación contemporánea y el pasado de la innovación hasta el retardo 1.

Para verificar si el proceso ARMA(1,1) es estacionario tiene que cumplir las siguientes propiedades:

$$E(Y_t) = \mu_t = \frac{\delta}{1 - \phi_1} \quad \text{para } \delta \neq 0; \quad |\phi_1| < 1 \quad \text{Estacionario} \quad \text{y} \quad |\theta_1| < 1 \quad \text{Invertible}$$

$$\begin{aligned} Var(Y_t) &= \gamma_0 = \frac{1 + \phi_1^2 - 2\phi_1\theta_1}{1 - \phi_1^2} \sigma_\varepsilon^2 \\ Cov(Y_t, Y_{t-h}) &= \gamma_h = \phi_1 \gamma_0 - \theta_1 \sigma_\varepsilon^2 \quad \text{si } h = 1; \quad \text{y } h \geq 2; \quad \gamma_h = \phi_1 \gamma_{h-1} \\ Cor(Y_t, Y_{t-h}) &= \rho_h = \phi_1 - \frac{\theta_1 \sigma_\varepsilon^2}{\gamma_0} \quad \text{para } h = 1; \quad \text{y } h \geq 2; \quad \gamma_h = \phi_1 \gamma_{h-1} \end{aligned}$$

Las característica más relevante del modelo ARMA(1,1) es que su autocorrelacion descienden gradualmente a cero, es decir decae geométricamente a partir de $h = 2, 3, \dots$, en lugar de detenerse bruscamente.

3.3.9 Proceso Lineal No Estacionario

Procesos Promedio Móvil Autorregresivo Integrado de Orden d, ARIMA(p,d,q)

Los modelos analizados hasta ahora se basan en el supuesto de estacionalidad de media y varianza constante en el tiempo y la covarianza invariante en el tiempo. Pero la mayoría de las series de tiempo no son estacionarias, por ende es necesario convertir el proceso estocástico no estacionario en un proceso estocástico integrado. Este proceso se puede modelar dentro de la clase de modelos ARMA(p,q). Cuando

esto ocurre entonces se habla de un modelo autorregresivo integrado de promedio móvil ARIMA(p,d,q).

Los modelo ARIMA(p,d,q) se lleva a cabo mediante los procesos Ar(p) y ARMA(p,q) que requieren que las series sean estacionarias.

Supongamos el siguiente modelo ARMA(p,q):

$$\phi_p(L)Y_t = \theta_q(L)\varepsilon_t$$

Donde el polinomio AR se puede factorizar en función de sus p raíces $L_1, L_2, \dots, L_p$.

$$\phi_p(L) = (1 - L_1^{-1}L)(1 - L_2^{-1}L) \cdots (1 - L_p^{-1}L)$$

Supongamos que $(p - 1)$ sus raíces son estacionarias (con módulo fuera del círculo unidad) y una de ellas es unitaria $L_1 = 1$, que es un tipo de modelo que genera comportamientos no estacionarios que no son explosivos, sino que las realizaciones se comportan de forma similar a lo largo del tiempo; por ello, modelaremos series no estacionarias mediante modelos ARMA(p,q), porque tienen al menos una raíz exactamente igual a la unidad, denominada raíz unitaria. Permitiendo que el polinomio AR, se puede reescribir como sigue:

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

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

Donde el polinomio $\varphi_{p-1}(L)$ resulta del producto de los $(p - 1)$ polinomio de orden 1 asociados a las raíces L_1 con módulo fuera del círculo unidad.

Sustituyendo en el modelo ARMA(p,q) se tiene que:

$$\varphi_{p-1}(L)(1 - L)Y_t = \theta_q(L)\varepsilon_t \longrightarrow \varphi_{p-1}(L) \triangle Y_t = \theta_q(L)\varepsilon_t$$

Donde el polinomio $\varphi_{p-1}(L)$ es estacionario porque todas sus raíces tienen modulo fuera del círculo unidad y el polinomio $\triangle = (1 - L)$ es el que recoge la raíz unitaria.

El dicho modelo representa el comportamiento de un proceso Y_t que no es estacionario porque tiene una raíz unitaria. A un proceso Y_t con estas características se le denomina proceso integrado de orden 1.

En general, el polinomio AR puede contener más de una raíz unitaria, por ejemplo, d , entonces se puede descomponer como:

$$\phi_p(L) = \varphi_{p-1}(L)(1 - L)^d$$

Y substituyendo de nuevo en el modelo ARMA(p,q), se tiene:

$$\varphi_{p-1}(L) \Delta^d Y_t = \theta_q(L) \varepsilon_t$$

Donde el polinomio $\varphi_{p-1}(L)$ es estacionario porque sus $(p - d)$ raíces tienen modulo fuera del circulo unidad, y el polinomio $\Delta^d = (1 - L)^d$, de orden d , contiene d raíces unitarias no estacionarias.

A un proceso Y_t con estas características se le denomina proceso integrado de orden d y se denota por $Y_t \sim I(d)$.

En general, si una serie Y_t es integrada de orden d , se puede representar por el siguiente modelo:

$$\phi_p(L) \Delta^d Y_t = \delta + \theta_q(L) \varepsilon_t$$

Donde p es el orden del polinomio autorregresivo estacionario, d es el orden de integración de la serie, es decir, el número de diferencia que hay que tomar a la serie para que sea estacionaria, y q es el orden del polinomio de medias móviles invertible.

Su expresión algebraica se define como:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_{p+d} Y_{t-p-d} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

Procedimiento de los Modelos ARIMA(p,d,q)

La construcción de los modelos ARIMA a través de la metodología Box-Jenkins se lleva a cabo de forma iterativa mediante un proceso en el que se pueden distinguir cuatro etapas:

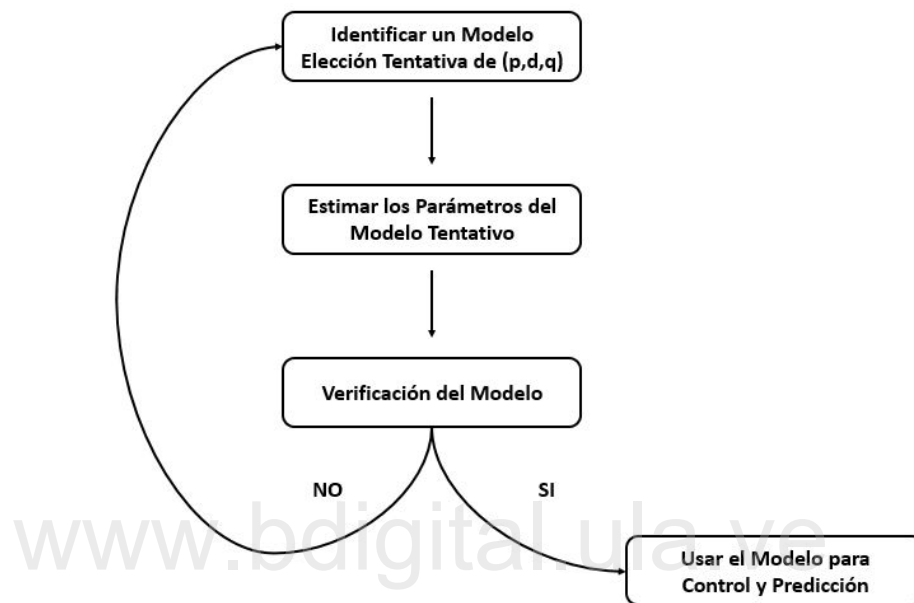


Figura 3.9: Metodología de Box-Jenkins

Fuente: Elaboración Propia.

Identificación:

El objetivo de esta etapa es seleccionar el modelo ARIMA(p,d,q) tentativo, apropiado para la serie. La identificación del modelo se lleva a cabo en dos fases:

- **1) Análisis de Estacionalidad:** Es el que determina las transformaciones que son necesarias aplicar para obtener una serie estacionaria en media y en varianza.

Prueba de Raíz Unitaria: La prueba de raíz unitaria se hace con el fin de verificar si la serie es estacionaria en varianza, por ello a la serie se le hará una prueba llamada Dickey-Fuller, donde dicha hipótesis nula afirma que la serie no es estacionaria, es decir, la existencia de raíz unitaria a un nivel de significación

$\alpha = 0.05$ si el valor muestral del estadístico t es menor que el valor crítico de las tablas de Dickey-Fuller, por lo tanto la idea es rechazar dicha hipótesis.

Transformación y Diferenciación: Si la anterior prueba no se rechaza, es decir, la serie no es estacionaria, entonces se realizará la denominada transformación de Box-Cox. Que trata de transformaciones que se le realizan a los datos para estabilizar la varianza de la serie; es una herramienta muy eficaz y de sencilla definición que busca darle una forma normal a los datos a través de:

$$Z_t = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda}, & \text{si } \lambda \neq 0 \\ \log(Y_t), & \text{si } \lambda = 0 \end{cases}$$

Donde λ se escoge minimizando la desviación estándar de los datos transformados, y se diferencia hasta lograr que la prueba de Dickey-Fuller se rechace, para continuar con el procedimiento. Dependiendo de cuantas veces se diferencie, será el número de diferencia d que hay que tomar para lograr que la serie sea estacionaria en media.

- **2) Elección de los órdenes p y q :** Una vez determinado el parámetro d , el siguiente paso es determinar el proceso estacionario ARMA(p, q) que la haya generado. El instrumento técnico básico para identificar un modelo ARMA(p, q) es la denominada función de autocorrelación simple y parcial (que mide el grado de correlación entre cada valor de la variable) que es de gran uso para la identificación de los órdenes de p y q del modelo ARMA adecuado, en el siguiente cuadro se resume los criterios que hacemos referencia.

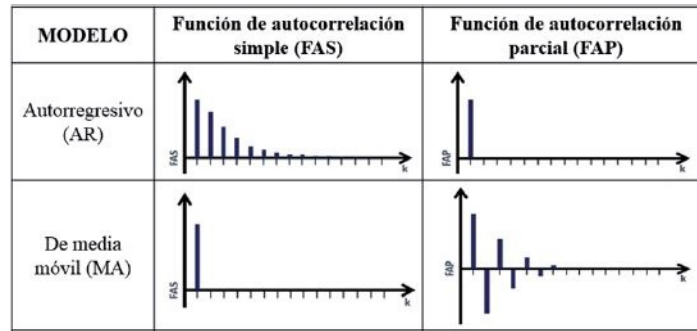


Figura 3.10: Curvas Típicas de las Funciones de Autocorrelación para Diferentes Modelos de Serie de Tiempo.

Fuente: researchgate.net

Estimación:

Una vez que se ha realizado la especificación del modelo tentativo y se han elegido los valores del orden autorregresivos p y el orden de las medias móviles q ; se procede a utilizar de forma eficiente los datos que realizarán inferencias sobre los parámetros condicionados $(\delta, \phi, \theta, \sigma)$, para que el modelo sea apropiado. Particularmente si los parámetros no satisfacen ciertas condiciones matemáticas, el modelo es rechazado. Estos parámetros se pueden estimar de forma consistente por los criterios de Máxima Verosimilitud o Mínimos Cuadrados. Ambos métodos de estimación se basan en el cálculo de las innovaciones ε_T a partir de valores de la serie estacionaria.

- **1) Criterio de Máxima Verosimilitud:** Se ha demostrado que los parámetros estimados a través de este criterio reflejan con gran exactitud las características presentes en los datos de la serie de tiempo, sin embargo, este método se ve limitado debido al gran tiempo que conlleva realizar la estimación. La función de verosimilitud se puede derivar a partir de la función de densidad conjunta de las innovaciones, $\varepsilon_1, \varepsilon_1, \dots, \varepsilon_T$, que bajo el supuesto de normalidad, es como sigue:

$$f(\varepsilon_1, \varepsilon_1, \dots, \varepsilon_T) \propto \sigma^{-T} \exp \left\{ \sum_{t=1}^T \frac{\varepsilon_t^2}{2\sigma^2} \right\}$$

- **2) Criterio de Mínimos Cuadrados:** Se ha demostrado que las estimaciones hechas por este criterio son muy cercanas a las del criterio de máxima

verosimilitud. La estimación de los parámetros del modelos ARMA seleccionado, se realiza a través de minimizar la suma de los cuadrados de los residuales. En general, estos estimados de los mínimos cuadrados deben obtenerse mediante un procedimiento no lineal de mínimos cuadrados. Es un algoritmo que encuentra el mínimo de la suma de la función de errores cuadrados, mediante un proceso iterativo.

Los residuales del modelo ARMA se define como:

$$\hat{\varepsilon}_t = Y_t - \hat{Y}_t$$

Donde *varepsilon* es el residual de la serie, Y_t es la serie original y $\hat{Y}_t$ la serie calculada con los parámetros estimados.

Entonces, la suma de los cuadrados de los residuales se define como:

$$SSR = \sum \hat{\varepsilon}_t^2$$

Esta última función es la que se debe minimizar, a través de la búsqueda de los parámetros $(\delta, \phi, \theta, \sigma)$ del modelo ARMA que se desea ajustar.

Para resolver el problema de estimación, las ecuaciones de máxima verosimilitud y la de mínimos cuadrados se puede expresar en función del conjunto de información y de los parámetros desconocidos del modelo. Para un modelo ARMA(p,q), la innovaciones se puede escribir como:

$$\varepsilon_T = Y_t - \delta - \sum_{t=1}^p \phi_t Y_{t-1} - \sum_{t=1}^q \theta_t \varepsilon_{t-1}$$

- **3) Criterios de información y Selección de Modelos:** Cuando se plantea la identificación de un modelo autorregresivo y de medias móviles (ARMA) a partir de la metodología de Box-Jenkins para predecir el comportamiento de una variable en un espacio temporal, se tiene que recurrir a la técnica de elección que exhiba al modelo óptimo. Para ejecutar esta técnica utilizaremos distintos criterios de información que presentaremos a continuación:

Criterio de Informacion Akaike (AIC): Es uno de los criterios más utilizados. El criterio de información Akaike no es una prueba de un modelo en el sentido de probar una hipótesis nula, se trata de un medio o herramienta para seleccionar el mejor modelo de un conjunto de modelos tentativos; el modelo preferido es aquel que tendrá el valor mínimo en el AIC. Su idea clave es de penalizar un exceso de parámetros ajustados, algo que no hace el test asintótico de la chi-cadrado. El AIC es un estimador muestral de la esperanza de log-verosimilitud, que viene dado por la expresión general.

$$AIC = 2K - 2\ln(L)$$

Donde L es el máximo valor de la función de verosimilitud para el modelo estimado y k es el número de parámetros independientes estimado dentro del modelo.

Criterio de Schwartz (BIC): Es una medida de bondad de ajuste de un modelo estadístico, que a menudo es utilizado como un criterio de selección de modelos entre un conjunto finito de modelos. Schwarz (1978) sugirió que el AIC podría no ser asintóticamente justificable, presentó un criterio de información alternativo a partir de un enfoque bayesiano, el BIC. Con este criterio se penaliza el número de parámetros con $\ln n$, en lugar de 2. Así:

$$BIC = 2(\ln(L)) + (\ln(n))k$$

Donde L es el máximo valor de la función de verosimilitud para el modelo estimado y k es el número de parámetros independientes estimados dentro del modelo, mientras que n es el tamaño muestral. El criterio BIC suele devolver modelos con menos parámetros; esto se debe que el segundo sumando de su definición penaliza más el número de parámetros que el del criterio AIC.

Validación del Modelo:

En la etapa de validación, diagnosis o chequeo se procede a evaluar el modelo para validar si se ajusta a los datos. Se tiene que tener en cuenta:

- **1) Análisis de Coeficientes Estimados:** Las estimaciones de los coeficientes del modelo tienen que ser significativas para cumplir las condiciones de estacionalidad e invertibilidad que deben satisfacer los parámetros del modelo. para ello, se calculan las raíces del polinomio autorregresivo, $\hat{\phi}(L) = 0$, y las raíces del polinomio de medias móviles, $\hat{\theta}(L) = 0$. Si una de estas raíces esta próxima a la unidad podría indicar falta de estacionalidad o invertibilidad. También es conveniente analizar la matriz de covarianza entre los parámetros estimados con el fin de detectar la posible presencia de una indicación de las presencia de factores comunes en el modelo.
- **2) Análisis de los Residuos:** Verificar si los residuos del modelo tienen un comportamiento similar a las innovaciones, es decir, si son ruido blanco. Las operaciones siguientes serán referidas a los residuos del modelo identificado y estimado.
 - Examinar el grafico temporal de la serie de residuos estandarizados. La presencia de cualquier tipo de tendencia suele indicar la conveniencia de aplicar una diferencia regular o estacional (según sea la tendencia). Si se observa que la dispersión de los residuos no es constante, probablemente la transformación de Box-Cox incluida en la serie no es la adecuada (o quizás las perturbaciones del modelo presentan heterocedasticidad condicional).
 - Examinar el Grafico de Q-Q plot de la serie de residuos y utilizar el estadístico de Jarque-Bera para contrastar la hipótesis de normalidad de las perturbaciones, donde dicha hipótesis nula afirma que la serie se distribuye de forma normal, a un nivel de significación $\alpha = 0.05$, si el valor muestral es mayor o igual que el valor de significancia ? no se rechaza la hipótesis nula y por lo tanto se podría decir que la serie proviene de una distribución normal.
 - Examinar el FAS y FAP residuales. Un amortiguamiento muy lento de la FAS indica que la serie de residuo no es estacionaria, por que conviene aplicar una diferencia (regular o estacionaria, según convenga) adicional a la serie

Y_t . Si el FAS y FAP presentan alguna configuración reconocible, añadir al modelo los operadores AR y MA correspondiente. Para determinar si las autocorrelaciones simples entre las perturbaciones del modelo son todas iguales a cero hasta cierto retardo h , es a través del estadístico de Ljung-Box.

Predicción:

Una vez estimado y validado el modelo ARIMA cumpliendo satisfactoriamente los criterios de evaluación establecidos, se plantea su utilización para conseguir la mejor predicción de los valores a futuro de una serie a partir de su propia historia, es decir, predecir los siguientes h valores de nuestra serie $Y_{t+1}, \dots, Y_{t+h}$. Formalmente, a esta predicción se le denomina predicción con origen T y horizonte h y se denota por $\hat{Y}_T(h)$.

- **1) Predictor Óptimo:** El predictor óptimo de Y_{t+h} viene dado por la esperanza condicionada al conjunto de información. Donde h es el tiempo de adelanto y T el origen de la predicción. Esto se puede expresar:

$$\hat{Y}_T(h) = E(Y_{T+h}/Y_T, Y_{T-1}, \dots)$$

Un predictor es óptimo cuando su valor esperado es igual al valor real de predicción condicionado a la información existente en el periodo T de la serie. El error cuadrático medio de un predictor arbitrario siempre es mayor que aquel cuyo valor coincide con la esperanza del valor real en el periodo que estamos considerando. Esta propiedad será fundamental para el posterior desarrollo de la predicción puntual.

- **2) Predicción Puntual:** Para realizar predicciones puntuales mediante un modelo ARIMA, se utiliza su expresión algebraica:

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_{p+d} Y_{t-p-d} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}$$

Para hacer predicciones hay que tener en cuenta dos supuestos iniciales:

- Los parámetros de las funciones presentadas son conocidos.
- Las perturbaciones aleatorias se conocen en el periodo muestral y tiene carácter de ruido blanco para los periodos de predicción.

Teniendo en cuenta ambos supuestos, se puede especificar un modelo ARIMA para definir el valor de predicción en función de la serie en p periodos anteriores y de los q errores cometidos al estimar la serie en los q periodos previos. Se plantea la ecuación de predicción de un periodo $(T + h)$ como:

$$\hat{Y}_T(h) = [\phi_1 \hat{Y}_T(h-1) + \phi_2 \hat{Y}_T(h-2) + \phi_3 \hat{Y}_T(h-3) + \cdots + \phi_{h+1} Y_{T-h} + \cdots + \phi_{p+d} Y_{T+h-p-d} + \delta - \theta_h \varepsilon_T + \theta_{h+1} \varepsilon_{T-1} - \cdots - \theta_q \varepsilon_{T-h-q}]$$

Donde $\hat{Y}_T(h)$ es el valor de predicción de la serie para el periodo $(T + 1)$, h es el tiempo de adelanto y T el origen de la predicción.

Los valores de predicción se calculan de forma secuencial, para hacer posible la aplicación de la parte autorregresiva del modelo en periodos distintos al primero de predicción, se toma el valor predicho en el inmediatamente anterior $\hat{Y}_T(h-1)$.

Por condición necesaria para estar ante el predictor óptimo (el valor esperado de la serie en el periodo de predicción es igual al predicho si es óptimo), se demuestra que las perturbaciones aleatorias empleadas en la predicción son solo la de periodos anteriores, es decir:

$$e_{T+j} \quad \forall \quad j \leq 0$$

Ya que se ha supuesto que en el periodo de predicción, las perturbaciones aleatorias tienen esperanza nula.

Para realizar sucesivas predicciones de $\hat{Y}_T(h)$ se tiene que contar con los valores $\varepsilon_T, \varepsilon_{T+1}, \cdots, \varepsilon_{T+1-q}$. Como tales se tomarán:

$$e_{T+j} = Y_{T-j} - \hat{Y}_{T+j+h}(h) \quad \forall \quad j \leq 0$$

Las predicciones de los modelos ARIMA son adaptativas y los resultados obtenidos para $(T + h)$, con la información disponible hasta el periodo T , serán las misma que se obtendría para ese periodo tomando como base informativa hasta $T - h$, y añadiéndole un término del error.

- **3) Predicción por intervalos:** En estadística, a la hora de predecir o estimar un valor, incluso más importante que el valor en sí, puede ser la aportación de un intervalo de confianza. Para su construcción, se utiliza la distribución muestral del error de predicción ε_T :

$$e_T = Y_{T+h} - \hat{Y}_T(h)$$

Sabiendo que Y_{T+h} se puede expresar como un $MA(\infty)$

$$Y_{T+h} = \delta + \varepsilon_{T+h} + w_1\varepsilon_{T+h-1} + \cdots + w_{h-1}\varepsilon_{T+1} + w_h\varepsilon_T + \cdots$$

$$\hat{Y}_t(h) = \delta + w_h\varepsilon_T + w_{h+1}\varepsilon_{T-1} + \cdots$$

$$e_T = Y_{T+h} - \hat{Y}_t(h) = \varepsilon_{T+h} + w_1\varepsilon_{T+h-1} + \cdots + w_{h-1}\varepsilon_{T+1}$$

$$\varepsilon_T \sim IIN(0, \sigma_\varepsilon^2)$$

$$e_T = (h) \sim N(0, Var(e_T(h)))$$

$$Var(e_T(h)) = \sigma_\varepsilon^2(1 + w_1^2 + w_2^2 + \cdots + w_{h-1}^2)$$

El intervalo de confianza de predicción para los modelos ARIMA con un nivel de confianza α será:

$$\hat{Y}_t(h) = \pm z_1 - \alpha \sqrt{\sigma_\varepsilon^2(1 + w_1^2 + w_2^2 + \cdots + w_{h-1}^2)}$$

Capítulo 4

Modelización de los Modelos Planteados

Para el estudio se dispone de un conjunto de datos correspondiente al precio de cierre de las criptomonedas seleccionadas. Caso: Bitcoin, Ethereum y Litecoin, extraído de la página [coinmarketcap](https://coinmarketcap.com) desde 1/1/2016 hasta 22/6/2021, para un total de 2000 observaciones por criptomoneda.

Las gráficas y los análisis estadísticos se ejecutó a través del software de R, mediante la interfaz de Rstudio, para la optimización del proceso.

4.1 Regresión Lineal Simple

4.1.1 Bitcoin

Se estima un modelo de regresión lineal simple a la criptomoneda Bitcoin, tomando como variable dependiente al precio de cierre diario e independiente al tiempo. Se ejecuta la presentación gráfica de ambas variables para intuir a la existencia de relación lineal, mediante la nube de puntos.

```
> attach(DatosCriptomonedas)
> library(ggplot2)
> ggplot(data = DatosCriptomonedas, mapping = aes(x = Tiempo, y =
```

```
Precio_Bitcoin)) + geom_point(color = "black", size = 2) + labs(title = "x = '
Tiempo', y = ' Precio') + theme_bw() + theme(plot.title = element_text(hjust = 0.5))
```

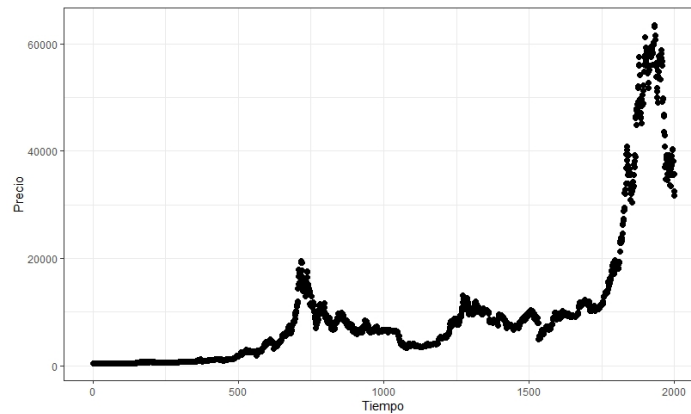


Figura 4.1.1: Diagrama de Dispersión del Bitcoin

Fuente: Elaboración Propia.

Una vez ejecutado el código en Rstudio para la ilustración del diagrama de dispersión del Bitcoin. Visualmente se puede asumir un cierto grado de relación lineal entre ambas variables, en este caso se modeliza dicha relación para ver si se ajusta a una recta.

Prueba de Correlación de Pearson

Para corroborar la existencia y qué tan fuerte es la relación lineal entre las variables se utiliza el coeficiente de correlación de Pearson, que indica lo siguiente:



Figura 4.1.2: Interpretación de los Valores del Coeficiente de Correlación de Pearson

Fuente: ResearchGate.net

Adicionalmente del valor que se obtenga del coeficiente, es necesario verificar su significancia. Solo si el p -value es significativo se puede aceptar que existe correlación y esta será de la magnitud que indique el coeficiente. Por muy cercano que sea el valor del coeficiente de correlación a $+1$ o -1 , si no es significativo, se ha de interpretar que la correlación de ambas variables es 0 ya que el valor observado se puede deber al azar.

```
> cor.test(x = DatosCriptomonedas$Tiempo,
y = DatosCriptomonedas$Precio_Bitcoin, method = "pearson")
```

```
Pearson's product-moment correlation
data: DatosCriptomonedas$Tiempo and DatosCriptomonedas$Precio_Bitcoin
t = 41.869, df = 1998, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.6595533 0.7062894
sample estimates:
cor
0.6836215
```

Figura 4.1.3: Prueba. Correlación de Pearson

Fuente: Elaboración Propia.

El gráfico de dispersión y el test de correlación muestran una relación lineal, de intensidad considerable ($r = 0.68$) y significativa (p -valor = $2.2e^{-16}$). Una vez verificado la existencia de relación lineal se procede al ajuste del modelo lineal, que permita predecir el precio del Bitcoin en función del tiempo.

Calidad de Ajuste del Modelo

Para estimar los parámetros de la recta nos centraremos en el modelo de mínimos cuadrados por ser el de más amplia aceptación.

```
> modelo_lineal_Bitcoin <- lm(Precio_Bitcoin ~ Tiempo,
DatosCRIPTOMonedas)
> modelo_lineal_Bitcoin
```

```
Call:
lm(formula = Precio_Bitcoin ~ Tiempo, data = DatosCRIPTOMonedas)

Coefficients:
(Intercept)      Tiempo
    -5242.73         14.86
```

Figura 4.1.4: Estimaciones del Modelo

Fuente: Elaboración Propia.

Al ejecutar la función `lm` en *Rstudio* se obtiene los parámetros de la ecuación de la recta de mínimos cuadrados que relaciona el precio del Bitcoin con el tiempo.

$$\hat{Y} = -5242,73 + 14,85 * (Tiempo)$$

Una vez obtenido los parámetros $\hat{\beta}_0$ y $\hat{\beta}_1$, es necesario la información detallada del modelo de regresión. Se puede generar aplicando la función `summary` a la variable que contiene los datos de la regresión, como se puede apreciar a continuación:

```
> summary(modelo_lineal_Bitcoin)
```

```

Call:
lm(formula = Precio_Bitcoin ~ Tiempo, data = DatosCriptomonedas)

Residuals:
    Min       1Q   Median       3Q      Max
-12576  -7102   -432    2640   40075

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -5242.7327   409.8545  -12.79  <2e-16 ***
Tiempo       14.8554     0.3548   41.87  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9161 on 1998 degrees of freedom
Multiple R-squared:  0.4673,    Adjusted R-squared:  0.4671
F-statistic: 1753 on 1 and 1998 DF,  p-value: < 2.2e-16

```

Figura 4.1.5: Sumario. Modelo de Regresión Lineal Simple del Bitcoin

Fuente: Elaboración Propia.

Este resultado es un amplio sumario de la regresión, donde se puede apreciar la información sobre los residuos (en el apartado *Residuals*), que se definen como la diferencia entre el verdadero valor de la variable dependiente y el valor que pronostica el modelo de regresión. Cuando más pequeños sean estos residuos mejor será el ajuste del modelo a los datos y más acertadas serán las predicciones que se realicen a partir de dicho modelo.

En la tabla *Coefficients* se visualiza los valores de los parámetros que aparecían en la salida por defecto junto a su error estándar. Cada parámetro aparece acompañado del valor de un estadístico t de *Student* y un p – *valor* que sirven para contrastar la significación del parámetro en cuestión, es decir, para resolver los siguientes contrastes de hipótesis:

$$\begin{aligned}
 H_0 &\equiv \beta_0 = 0 \quad vs \quad H_1 \equiv \beta_0 \neq 0 \\
 H_1 &\equiv \beta_0 = 0 \quad vs \quad H_1 \equiv \beta_1 \neq 0
 \end{aligned}$$

Lo que se pretende mediante estos contrastes, es determinar si los efectos de la constante y de la variable independiente son realmente importantes para explicar la variable dependiente o si, por el contrario, pueden considerarse nulos.

En el modelo los p – valores, ambos son $2.2e^{-16}$ menores que 0.05. Así, considerando un nivel de significación del 5%, se rechaza la hipótesis nula en ambos contrastes, asumiendo que ambos parámetros son significativamente distintos de 0.

En la parte final de la salida, encontramos el valor de R^2 (*MultipleR – squared*) y de R_{aj}^2 (*AdjustedR – squared*), que son indicadores de la bondad de ajuste del modelo a los datos. R^2 oscila entre 0 y 1, de manera que los valores de R^2 próximos a 1 indican un buen ajuste del modelo lineal a los datos. Por otro lado, R_{aj}^2 es similar a R^2 , pero penaliza la introducción en el modelo de variables independientes poco relevantes a la hora de explicar la variable dependiente. Por tanto, en el modelo el $R^2 = 0.4673$ y $R_{aj}^2 = 0.4671$, con este resultado se puede concluir que el modelo lineal se ajusta de forma aceptable a los datos.

La última línea de la salida incluye un estadístico F , y el p – valor correspondiente que se utilizan para resolver el siguiente contraste:

$$H_0 \equiv \beta_i = 0 \quad \forall_i$$

$$H_1 \equiv \beta_i \neq 0 \quad \text{para al menos un } i$$

Mediante este contraste se comprueba de forma global, si el modelo lineal es apropiado para modelizar los datos. En nuestro caso, el p – valor = $2.2e^{-16}$ asociado a este contraste es inferior a 0.05 por lo que, al 5% de significación, se rechaza la hipótesis nula y afirmar que, efectivamente, el modelo lineal es adecuado para nuestro conjunto de datos.

Se procede a dibujar un vector de predicciones basado en el propio conjunto de entrenamiento sobre el diagrama de dispersión mediante la orden *ggplot* en *RStudio*. De este modo se puede visualizar la distancia existente entre los valores observados y los valores que el modelo pronostica.

```
> ggplot(data = DatosCriptomonedas, mapping = aes(x = Tiempo, y =
Precio_Bitcoin)) +
geom_point(color = "black", size = 2) +
```

```
labs(title = "", x = 'Tiempo', y = 'Precio') +  
geom_smooth(method = "lm", se = TRUE, color = "red") + theme_bw() +  
theme(plot.title = element_text(hjust = 0.5))
```

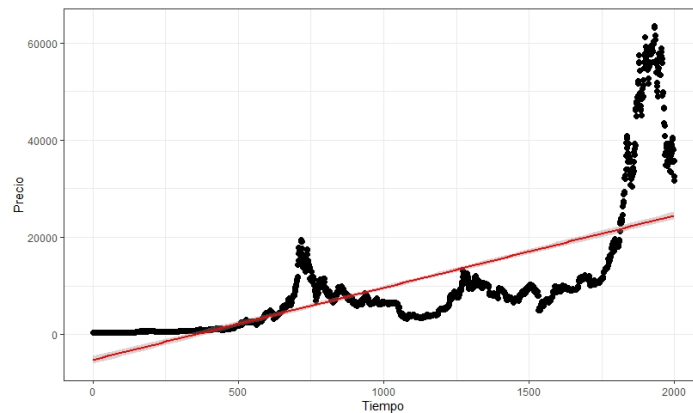


Figura 4.1.6: Recta de Regresión Lineal Simple del Bitcoin

Fuente: Elaboración Propia.

Pronósticos de Regresión lineal Simple del Bitcoin

Gráficamente se puede visualizar la recta que pronostica el modelo, para obtener predicciones será necesario tomar 30 observaciones posteriores del tiempo, ejecutando la recta de mínimos cuadrados para la obtención de las predicciones.

Nº	Fecha	Precios Reales	Pronósticos	Error
1	23/6/2021	33723,03	24472,12	38%
2	24/6/2021	34662,44	24486,97	42%
3	25/6/2021	31637,78	24501,82	29%
4	26/6/2021	32186,28	24516,67	31%
5	27/6/2021	34649,64	24531,52	41%
6	28/6/2021	34434,34	24546,37	40%
7	29/6/2021	35867,78	24561,22	46%
8	30/6/2021	35040,84	24576,07	43%
9	1/7/2021	33572,12	24590,92	37%
10	2/7/2021	33897,05	24605,77	38%
11	3/7/2021	34668,55	24620,62	41%
12	4/7/2021	35287,78	24635,47	43%
13	5/7/2021	33746,00	24650,32	37%
14	6/7/2021	34235,19	24665,17	39%
15	7/7/2021	33855,33	24680,02	37%
16	8/7/2021	32877,37	24694,87	33%
17	9/7/2021	33798,01	24709,72	37%
18	10/7/2021	33520,52	24724,57	36%
19	11/7/2021	34240,19	24739,42	38%
20	12/7/2021	33155,85	24754,27	34%
21	13/7/2021	32702,03	24769,12	32%
22	14/7/2021	32822,35	24783,97	32%
23	15/7/2021	31780,73	24798,82	28%
24	16/7/2021	31421,54	24813,67	27%
25	17/7/2021	31533,07	24828,52	27%
26	18/7/2021	31796,81	24843,37	28%
27	19/7/2021	30817,83	24858,22	24%
28	20/7/2021	29807,35	24873,07	20%
29	21/7/2021	32110,69	24887,92	29%
30	22/7/2021	32313,11	24902,77	30%

Tabla 4.1: Comparación de los Resultados del Pronóstico de Regresión Lineal Simple del Bitcoin

Tras realizar las predicciones se utilizó la desviación media absoluta para medir la dispersión del error total, resultando un 35% en comparación con los precios reales del Bitcoin, en un período de 30 días posteriores a las observaciones estudiadas. Se concluye que la regresión lineal simple para predecir el precio diario del Bitcoin es eficientemente aceptable considerando las altas fluctuaciones que presenta este criptoactivo en el precio.

4.1.2 Ethereum

Para la regresión lineal simple de la criptomoneda Ethereum, se estima un modelo para la obtención de predicciones, tomando el precio de cierre diario como variable dependiente y al tiempo como independiente, ejecutando los comandos en *Rstudio* similares al apartado anterior del modelo de regresión lineal simple del Bitcoin para la realización de los análisis.

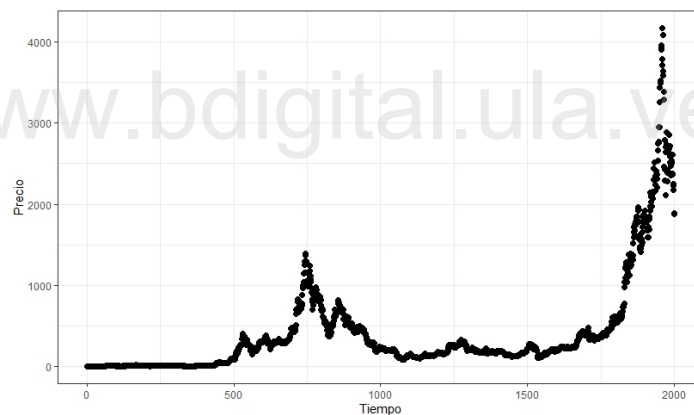


Figura 4.1.7: Diagrama de Dispersión del Ethereum

Fuente: Elaboración Propia.

En el diagrama de dispersión se observa que los datos tienden agruparse linealmente, para la verificación de dicha relación entre las variables, es a través de la prueba de correlación de Pearson.

```

Pearson's product-moment correlation

data: DatosCRIPTOMonedas$Tiempo and DatosCRIPTOMonedas$Precio_Ethereum
t = 28.308, df = 1998, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.5029971 0.5655994
sample estimates:
      cor
0.5350323

```

Figura 4.1.8: Prueba. Correlación de Pearson

Fuente: Elaboración Propia.

El valor arrojado por la prueba de correlación de Pearson, verifica que existe una relación lineal positiva moderada entre las variables de ($r = 0.53$) con significancia de ($p - valor = 2.2e^{-16}$) inferior del 5%, corroborando la relación existente de ambas variables, permitiendo proceder a la estimación del modelo que intuirá a la ayuda de los pronósticos del precio del Ethereum.

```

Call:
lm(formula = Precio_Ethereum ~ Tiempo, data = DatosCRIPTOMonedas)

Residuals:
    Min       1Q   Median       3Q      Max
-586.8 -357.7  -38.6   126.5  3237.8

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -155.06096    22.63164   -6.852  9.7e-12 ***
Tiempo        0.55461     0.01959   28.308 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 505.9 on 1998 degrees of freedom
Multiple R-squared:  0.2863,    Adjusted R-squared:  0.2859
F-statistic: 801.3 on 1 and 1998 DF,  p-value: < 2.2e-16

```

Figura 4.1.9: Sumario. Modelo de Regresión Lineal Simple del Ethereum

Fuente: Elaboración Propia.

Una vez estimado el modelo de mínimos cuadrados se aprecia la calidad describiendo el sumario de la regresión. Observando en la tabla de los coeficientes se obtienen los parámetros de $\hat{\beta}_0$ y $\hat{\beta}_1$ ambos con los $p - valores$ inferiores al nivel de significancia del 5%, siendo significativos y potencialmente importantes para la ecuación de la recta que se describe a continuación:

$$\hat{Y} = -155,06 + 0,55 * (Tiempo)$$

Otro dato importante son los indicadores de la bondad de ajuste, se pueden visualizar en el apartado inferior del sumario, que indica el valor de $R^2 = 0.2863$ y $R_{aj}^2 = 0.2859$, con ellos podemos afirmar que la variabilidad explicada por el precio del Ethereum, es de 28% predicha por el tiempo.

Para aceptar que el modelo lineal es adecuado, se visualiza en la parte inferior derecha del sumario en el $p - valor$, que indica un valor de $2.2e^{-16}$ muy por debajo del contraste de significancia del 5%, rechazando la hipótesis nula, afirmando que el modelo es apropiado.

Una vez, aceptado el modelo lineal del conjunto de datos estimados de la criptomoneda Ethereum, se procede a trazar un vector de predicciones basado en el propio conjunto de entrenamiento sobre el diagrama de dispersión para la visualización de las distancias existentes entre el vector y el conjunto de datos.

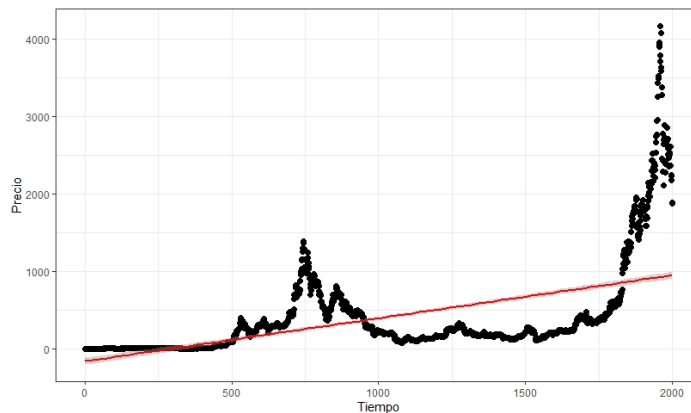


Figura 4.1.10: Recta de Regresión Lineal Simple del Ethereum

Fuente: Elaboración Propia.

Se ejecuta las predicciones a través de la ecuación de la recta de mínimos cuadrados, tomando 30 observaciones posteriores del tiempo para la obtención de los pronósticos.

Nº	Fecha	Precios Reales	Pronósticos	Error
1	23/6/2021	1989,74	954,71	108%
2	24/6/2021	1988,46	955,27	108%
3	25/6/2021	1813,22	955,82	90%
4	26/6/2021	1829,24	956,38	91%
5	27/6/2021	1978,89	956,93	107%
6	28/6/2021	2079,66	957,49	117%
7	29/6/2021	2160,77	958,04	126%
8	30/6/2021	2274,55	958,60	137%
9	1/7/2021	2113,61	959,15	120%
10	2/7/2021	2150,04	959,71	124%
11	3/7/2021	2226,11	960,26	132%
12	4/7/2021	2321,72	960,81	142%
13	5/7/2021	2198,58	961,37	129%
14	6/7/2021	2324,68	961,92	142%
15	7/7/2021	2315,16	962,48	141%
16	8/7/2021	2120,03	963,03	120%
17	9/7/2021	2146,69	963,59	123%
18	10/7/2021	2111,4	964,14	119%
19	11/7/2021	2139,66	964,70	122%
20	12/7/2021	2036,72	965,25	111%
21	13/7/2021	1940,08	965,81	101%
22	14/7/2021	1994,33	966,36	106%
23	15/7/2021	1911,18	966,92	98%
24	16/7/2021	1880,38	967,47	94%
25	17/7/2021	1898,83	968,02	96%
26	18/7/2021	1895,55	968,58	96%
27	19/7/2021	1817,3	969,13	88%
28	20/7/2021	1787,51	969,69	84%
29	21/7/2021	1990,97	970,24	105%
30	22/7/2021	2025,2	970,80	109%

Tabla 4.2: Comparación de los Resultados del Pronósticos de Regresión Lineal Simple del Ethereum

Tras la realización de los pronósticos se comparan con los precios reales del Ethereum, para permitir un enfoque prominente sobre qué tan preciso es el modelo estimado para la realización de las predicciones, resultando con un porcentaje de error medio de 113%.

4.1.3 Litecoin

En el modelo de regresión lineal simple para la criptomoneda Litecoin se toma el precio de cierre diario como variable dependiente y al tiempo como independiente. Se inicia generando un gráfico de dispersión con el fin de tener una idea visual de la relación entre las variables.

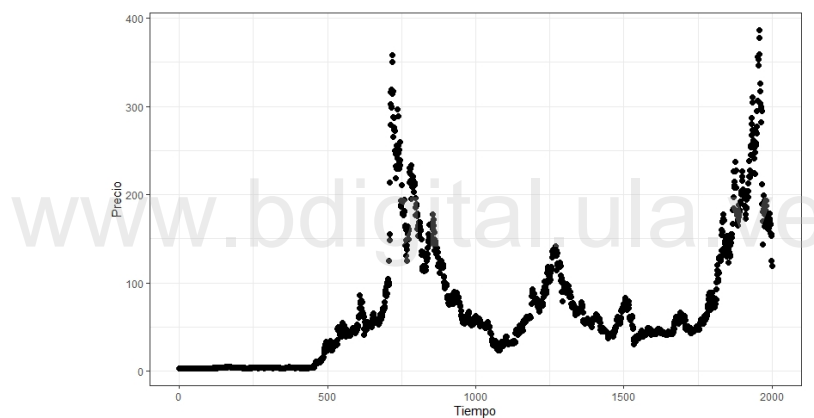


Figura 4.1.11: Diagrama de Dispersión del Litecoin

Fuente: Elaboración Propia.

Se puede visualizar en el gráfico de dispersión las tendencias en el conjunto de puntos, para la verificación de la relación entre las variables se respalda a través de la prueba de correlación de Pearson.

```

Pearson's product-moment correlation

data: DatosCriptomonedas$Tiempo and DatosCriptomonedas$Precio_Litecoin
t = 26.281, df = 1998, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.4735304 0.5387051
sample estimates:
      cor
0.5068417

```

Figura 4.1.12: Prueba. Correlación de Pearson

Fuente: Elaboración Propia.

Efectivamente la prueba indica una correlación débil de ($r = 0.50$) pero significativa, permitiendo aplicar una regresión sobre los datos para la obtención de la línea recta que mejor ajuste la relación existente entre la variable independiente y la variable dependiente.

```

Call:
lm(formula = Precio_Litecoin ~ Tiempo, data = DatosCriptomonedas)

Residuals:
    Min       1Q   Median       3Q      Max
-70.37 -32.91 -14.74  11.94 305.51

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.129064   2.613891   3.875  0.00011 ***
Tiempo       0.059470   0.002263  26.281 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 58.43 on 1998 degrees of freedom
Multiple R-squared:  0.2569,    Adjusted R-squared:  0.2565
F-statistic: 690.7 on 1 and 1998 DF,  p-value: < 2.2e-16

```

Figura 4.1.13: Sumario. Modelo de Regresión Lineal Simple del Ethereum

Fuente: Elaboración Propia.

Al realizar el sumario para apreciar la calidad del ajuste del modelo, vemos en el apartado de los coeficientes de determinación los valores de $\beta_0 = 10,129064$ y $\beta_1 = 0,05947$, siendo ambos significativo con un el p – *valor* inferior al 5%. Por lo tanto, se pueden incluir en la ecuación de la recta de mínimos cuadrado.

$$\hat{Y} = 10,12 + 0,05 * (Tiempo)$$

En la parte inferior del sumario se aprecian los valores de $R^2 = 0.2569$ y el $R_{aj}^2 = 0.2565$ afirmando una variabilidad del 25% predicha por el tiempo, en el mismo apartado se obtiene el p -valor, que comprueba de forma global, si el modelo lineal es apropiado para modelizar los datos, siendo efectivamente significativo, comprobando que, efectivamente el modelo es adecuado.

Una vez construido el modelo, se crea un vector de predicciones basado en el propio conjunto de entrenamiento, con este vector se visualiza la distancia existente entre los valores observados.

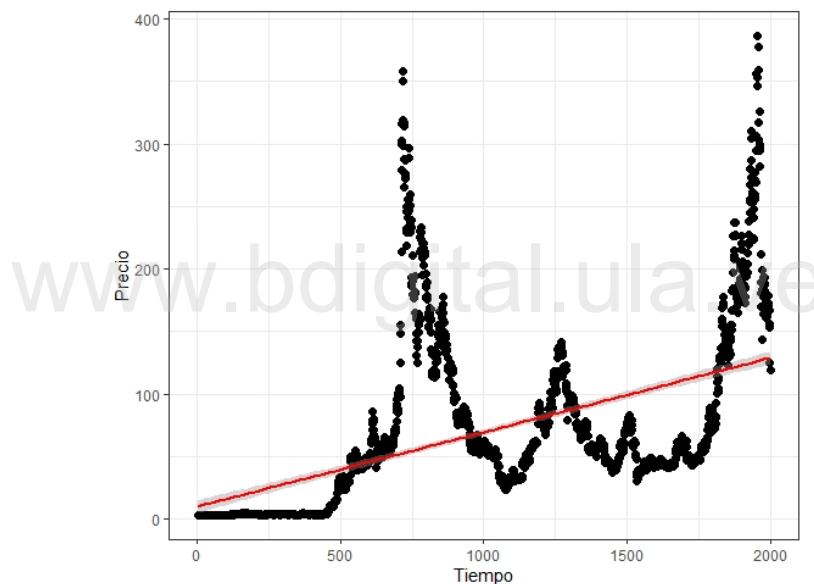


Figura 4.1.14: Recta de Regresión Lineal Simple del Ethereum

Fuente: Elaboración Propia.

Por último, al tener el modelo regresor listo, se generan los pronósticos de 30 observaciones posteriores a las utilizadas del tiempo.

Nº	Fecha	Precios Reales	Pronósticos	Error
1	23/6/2021	128,67	129,13	0%
2	24/6/2021	134,48	129,19	4%
3	25/6/2021	125,69	129,25	3%
4	26/6/2021	126,66	129,31	2%
5	27/6/2021	132,35	129,37	2%
6	28/6/2021	137,37	129,43	6%
7	29/6/2021	143,7	129,49	11%
8	30/6/2021	144,14	129,54	11%
9	1/7/2021	137,58	129,60	6%
10	2/7/2021	136,94	129,66	6%
11	3/7/2021	140,28	129,72	8%
12	4/7/2021	144,91	129,78	12%
13	5/7/2021	138,07	129,84	6%
14	6/7/2021	138,99	129,90	7%
15	7/7/2021	137,37	129,96	6%
16	8/7/2021	132,31	130,02	2%
17	9/7/2021	134,57	130,08	3%
18	10/7/2021	133,99	130,14	3%
19	11/7/2021	134,25	130,20	3%
20	12/7/2021	133,54	130,26	3%
21	13/7/2021	131,32	130,32	1%
22	14/7/2021	131,06	130,38	1%
23	15/7/2021	125,49	130,44	4%
24	16/7/2021	121,04	130,50	7%
25	17/7/2021	120,03	130,56	8%
26	18/7/2021	119,51	130,62	9%
27	19/7/2021	113,16	130,67	13%
28	20/7/2021	107,4	130,73	18%
29	21/7/2021	117,46	130,79	10%
30	22/7/2021	120,74	130,85	8%

Tabla 4.3: Comparación de los Resultados del Pronósticos de Regresión Lineal Simple del Litecoin

Los resultados de los pronósticos son bastante similares en comparación a los precios reales, arrojando un error medio de 6% en un periodo de 30 días.

4.2 Series de Tiempo

A continuación se aprecian los análisis de series de tiempo de las criptomonedas seleccionadas. Caso: Bitcoin, Ethereum y Litecoin. Se ajustaron modelos ARIMA (modelos autorregresivos integrados de medias móviles), conocidos también como metodologías Box-Jenkins. La metodología indicada consiste en cuatro pasos: la identificación, la estimación, la validación y el pronóstico.

4.2.1 Bitcoin

Identificación

El primer paso es la identificación; Por ende, es importante observar el comportamiento de la serie a través de la siguiente gráfica:

```
> attach(DatosCriptomonedas)
> Bit <- ts(data = Precio ~ Bitcoin, start = c(2016, 1), frequency = 365)
> plot(Bit, xlab = "Tiempo", ylab = "Precio")
```

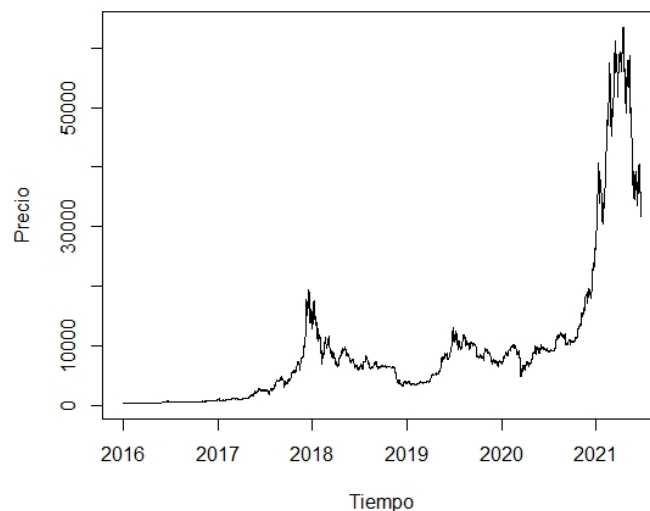


Figura 4.2.1: Gráfica. Serie Original del Bitcoin

Fuente: Elaboración Propia.

En el gráfico se presenta el comportamiento de los precios de cierre de la criptomoneda Bitcoin del 1/1/2016 hasta 22/6/2021. Se observa que el primer trimestre del año 2017 la serie no muestra mayores fluctuaciones. Sin embargo a partir de este punto se comienza a apreciar una variabilidad en los precios alcanzando uno de sus máximos apogeos en las navidades del 2017, el intercambio de estas divisas electrónicas se popularizaron, el Bitcoin llegó a cotizarse por encima de los 20.000 dólares. Apenas unas semanas después, ya en el año 2018, el Bitcoin comienza a caer en picada. No es hasta el segundo trimestre del año 2019, que comienza ascender manteniendo una fluctuación constante alrededor de los 8.000 dólares, en julio del año 2020, despliega una subida exponencial llegando a cotizarse en 63.503,46 dólares el 13 de abril del 2021, uno de los puntos máximos de la serie. Se atribuye este incremento a los temores por la inflación en los Estados Unidos.

- **Prueba de Estacionariedad de la Serie del Bitcoin**

Para corroborar que la serie es estacionaria, se verifica a través de los correlograma de funciones de autocorrelación y la prueba de raíz unitaria:

```
> acf(ts(Bit, frequency = 1), 200)
```

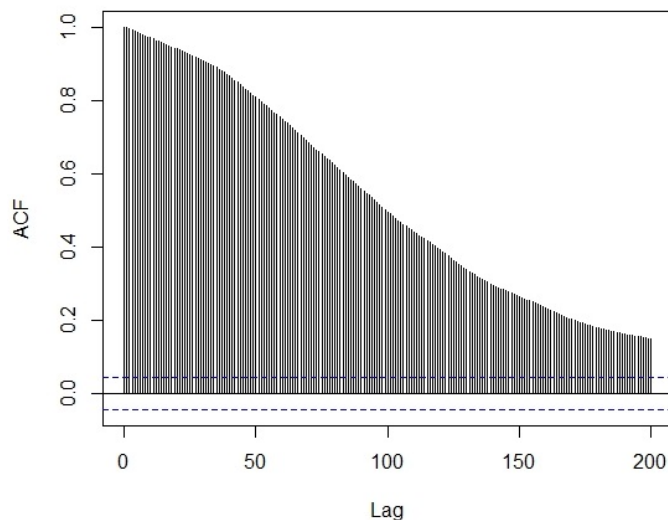


Figura 4.2.2: Gráfica. Función de Autocorrelación Simple del Bitcoin

Fuente: Elaboración Propia.

La gráfica de la función de autocorrelación simple ilustra un desvanecimiento lento indicando la no estacionariedad en la serie. Se verifica mediante la prueba de raíz unitaria a través del Test de Dickey-Fuller.

El Test de Dickey-Fuller permite saber si hay presencia significativa de tendencia en las series temporales de las variables mediante un contraste de hipótesis.

Hipótesis nula: $H_0 : \alpha = 0$ (es decir, existe una raíz unitaria, la serie de tiempo es no estacionaria o tiene tendencia estocástica).

Hipótesis alternativa: $H_A : \alpha < 0$ (es decir, la serie de tiempo es estacionaria, posiblemente alrededor de una tendencia determinista).

```
> adf.test(Bit)
```

```

Augmented Dickey-Fuller Test
data: Bit
Dickey-Fuller = -1.8616, Lag order = 12, p-value = 0.6369
alternative hypothesis: stationary

```

Figura 4.2.3: Prueba. Test de Dickey-Fuller, Serie Original del Bitcoin

Fuente: Elaboración Propia.

Tras realizar la prueba del Test Dickey-Fuller, obtenemos un $p - valor = 0.6369$. Como el $p - valor > 0.05$ no se rechaza H_0 . Se concluye que la serie es no estacionaria.

- **Diferenciación de la serie Original del Bitcoin**

Al resultar que la serie es no estacionaria, se aplica diferenciación de orden 1 para eliminar las variaciones irregulares de la serie.

```

> Bit1 = diff(Bit)
> plot(Bit1, xlab = "Tiempo", ylab = "Precio")

```

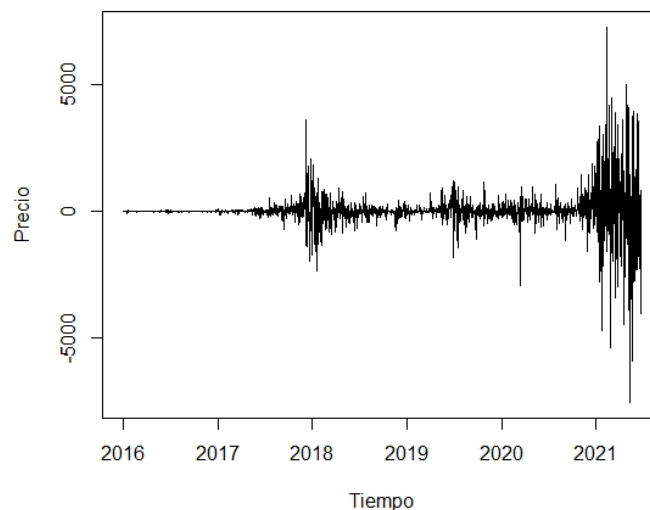


Figura 4.2.4: Gráfica. Serie con Diferenciación de Orden Uno (1) del Bitcoin

Fuente: Elaboración Propia.

Al aplicar la primera diferencia se observa que la serie tiene un patrón oscilante alrededor de 0, sin una tendencia fuerte visible, es decir, se estabiliza en torno a un valor medio. Para corroborar que la serie es efectivamente un proceso homogéneo, se procede de nuevo el test de Dickey-Fuller.

```
> adf.test(Bit1)
```

```
Augmented Dickey-Fuller Test  
data: Bit1  
Dickey-Fuller = -12.037, Lag order = 12, p-value = 0.01  
alternative hypothesis: stationary
```

Figura 4.2.5: Prueba. Test de Dickey-Fuller, Serie Diferenciada del Bitcoin

Fuente: Elaboración Propia.

El p -valor obtenido de la prueba es de 0.01. Como el p -valor < 0.05 , rechazamos H_0 . Con ello, podemos concluir que nuestra serie tiene estacionariedad en la primera diferenciación de los términos.

- **Funciones de Autocorrelación Simple y Parcial de la Serie Diferenciada del Bitcoin**

A continuación, los correlogramas de las funciones de autocorrelación simple y parcial de la serie diferenciada. Serán los informantes sobre la elección del orden del modelo autoregresivo (p) y el orden de medias móviles (q) para el modelo.

```
> acf(ts(Bit1, frequency = 1))
```

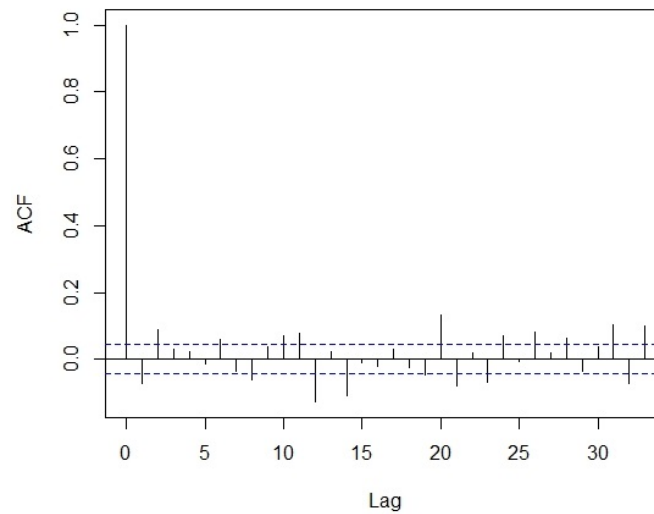


Figura 4.2.6: Gráfica. Función de Autocorrelación Simple

Fuente: Elaboración Propia.

Se observa en la función de autocorrelación simple de la serie diferenciada del Bitcoin, el primer pico se sale notoriamente de las bandas de significancia que se encuentra en el rezago 0, y será un posible valor que tome (q) para el modelo ARIMA.

```
> pacf(ts(Bit1, frequency = 1))
```

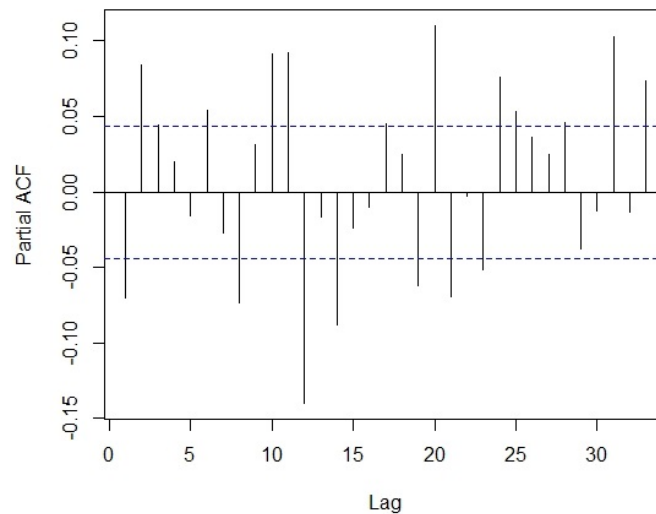


Figura 4.2.7: Gráfica. Función de Autocorrelación Parcial

Fuente: Elaboración Propia.

En la función de autocorrelación parcial varios rezagos se salen de las bandas de significancia, estos serán tentativos para el valor que tomará (p) .

Estimación

Se estima un conjunto de modelos tentativos de las diversas combinaciones de los parámetros (p) y (q) , tomando como referencia los correlograma, para permitir optar por el modelo ARIMA que mejor se adapte a los datos.

- **Modelo Tentativo del Bitcoin ARIMA(0,1,0)**

```
> Arima010 <- arima(Bit, c(0, 1, 0))
```

```
> Arima010
```

```
Call:
arima(x = Bit, order = c(0, 1, 0))

sigma^2 estimated as 503999: log likelihood = -15960.22, aic = 31922.45
```

Figura 4.2.8: Resumen. Modelo Tentativo del Bitcoin ARIMA(0,1,0)

Fuente: Elaboración Propia.

- **Modelo Tentativo del Bitcoin ARIMA(1,1,0)**

```
> Arima110 <- arima(Bit, c(1, 1, 0))
> Arima110
```

```
Call:
arima(x = Bit, order = c(1, 1, 0))

Coefficients:
          ar1
        -0.0696
s.e.      0.0223

sigma^2 estimated as 501554:  log likelihood = -15955.36,  aic = 31914.73
```

Figura 4.2.9: Resumen. Modelo Tentativo del Bitcoin ARIMA(1,1,0)

Fuente: Elaboración Propia.

- **Modelo Tentativo del Bitcoin ARIMA(2,1,0)**

```
> Arima210 <- arima(Bit, c(2, 1, 0))
> Arima210
```

```
Call:
arima(x = Bit, order = c(2, 1, 0))

Coefficients:
          ar1      ar2
        -0.0639    0.0862
s.e.      0.0223    0.0225

sigma^2 estimated as 497878:  log likelihood = -15948.02,  aic = 31902.04
```

Figura 4.2.10: Resumen. Modelo Tentativo del Bitcoin ARIMA(2,1,0)

Fuente: Elaboración Propia.

- **Modelo Tentativo del Bitcoin ARIMA(0,1,1)**

```
> Arima011 <- arima(Bit, c(0, 1, 1))
> Arima011
```

```
Call:
arima(x = Bit, order = c(0, 1, 1))

Coefficients:
      ma1
    -0.0590
s.e.    0.0204

sigma^2 estimated as 501925:  log likelihood = -15956.1,  aic = 31916.2
```

Figura 4.2.11: Resumen. Modelo Tentativo del Bitcoin ARIMA(0,1,1)

Fuente: Elaboración Propia.

- **Modelo Tentativo del Bitcoin ARIMA(1,1,1)**

```
> Arima111 <- -arima(Bit, c(1, 1, 1))
```

```
> Arima111
```

```
Call:
arima(x = Bit, order = c(1, 1, 1))

Coefficients:
      ar1      ma1
    -0.5797    0.5046
s.e.    0.1548    0.1632

sigma^2 estimated as 499749:  log likelihood = -15951.76,  aic = 31909.53
```

Figura 4.2.12: Resumen. Modelo Tentativo del Bitcoin ARIMA(1,1,1)

Fuente: Elaboración Propia.

- **Modelo Tentativo del Bitcoin ARIMA(2,1,1)**

```
> Arima211 <- -arima(Bit, c(2, 1, 1))
```

```
> Arima211
```

```
Call:
arima(x = Bit, order = c(2, 1, 1))

Coefficients:
      ar1      ar2      ma1
    0.4032    0.1201   -0.4716
s.e.    0.1693    0.0226    0.1702

sigma^2 estimated as 496708:  log likelihood = -15945.67,  aic = 31899.34
```

Figura 4.2.13: Resumen. Modelo Tentativo del Bitcoin ARIMA(2,1,1)

Fuente: Elaboración Propia.

- **Modelo Tentativo del Bitcoin ARIMA(0,1,2)**

```
> Arima012 <- arima(Bit, c(0, 1, 2))
> Arima012
```

```
Call:
arima(x = Bit, order = c(0, 1, 2))

Coefficients:
          ma1      ma2
      -0.0710   0.0941
s.e.    0.0224   0.0224

sigma^2 estimated as 497518:  log likelihood = -15947.3,  aic = 31900.59
```

Figura 4.2.14: Resumen. Modelo Tentativo del Bitcoin ARIMA(0,1,2)

Fuente: Elaboración Propia.

- **Modelo Tentativo del Bitcoin ARIMA(1,1,2)**

```
> Arima112 <- arima(Bit, c(1, 1, 2))
> Arima112
```

```
Call:
arima(x = Bit, order = c(1, 1, 2))

Coefficients:
      ar1      ma1      ma2
    0.4480  -0.5161   0.1191
s.e.   0.1943   0.1937   0.0224

sigma^2 estimated as 496739:  log likelihood = -15945.73,  aic = 31899.47
```

Figura 4.2.15: Resumen. Modelo Tentativo del Bitcoin ARIMA(1,1,2)

Fuente: Elaboración Propia.

- **Modelo Tentativo del Bitcoin ARIMA(2,1,2)**

```
> Arima212 <- arima(Bit, c(2, 1, 2))
> Arima212
```

```
Call:
arima(x = Bit, order = c(2, 1, 2))

Coefficients:
      ar1      ar2      ma1      ma2
    0.4174   0.0749  -0.4861   0.0458
s.e.   0.1882   0.1810   0.1886   0.1807

sigma^2 estimated as 496694:  log likelihood = -15945.64,  aic = 31901.29
```

Figura 4.2.16: Resumen. Modelo Tentativo del Bitcoin ARIMA(2,1,2)

Fuente: Elaboración Propia.

- Elección del Modelo ARIMA para el Bitcoin

Para elegir el mejor modelo ARIMA del Bitcoin, se tiene en cuenta el criterio de información Akaike (AIC), donde el modelo con menor AIC será representado como el mejor modelo.

N°	Modelos	AIC
1	ARIMA(0,1,0)	31.922,45
2	ARIMA(1,1,0)	31.914,73
3	ARIMA(2,1,0)	31.902,04
4	ARIMA(0,1,1)	31.916,20
5	ARIMA(1,1,1)	31.909,53
6	ARIMA(2,1,1)	31.899,34
7	ARIMA(0,1,2)	31.900,59
8	ARIMA(1,1,2)	31.899,47
9	ARIMA(2,1,2)	31.901,29

Tabla 4.4: Comparación de los Resultados del AIC

Según el criterio de información Akaike (AIC), el mejor modelo es el ARIMA(2,1,1), con un valor menor de AIC entre el conjunto de modelos estimados.

Validación

Después de estimar el modelo se analiza sus residuales para comprobar si el modelo ARIMA(2,1,1), se ajusta o no de forma adecuada a los datos observados.

```
> tsdiag(Arima211)
```

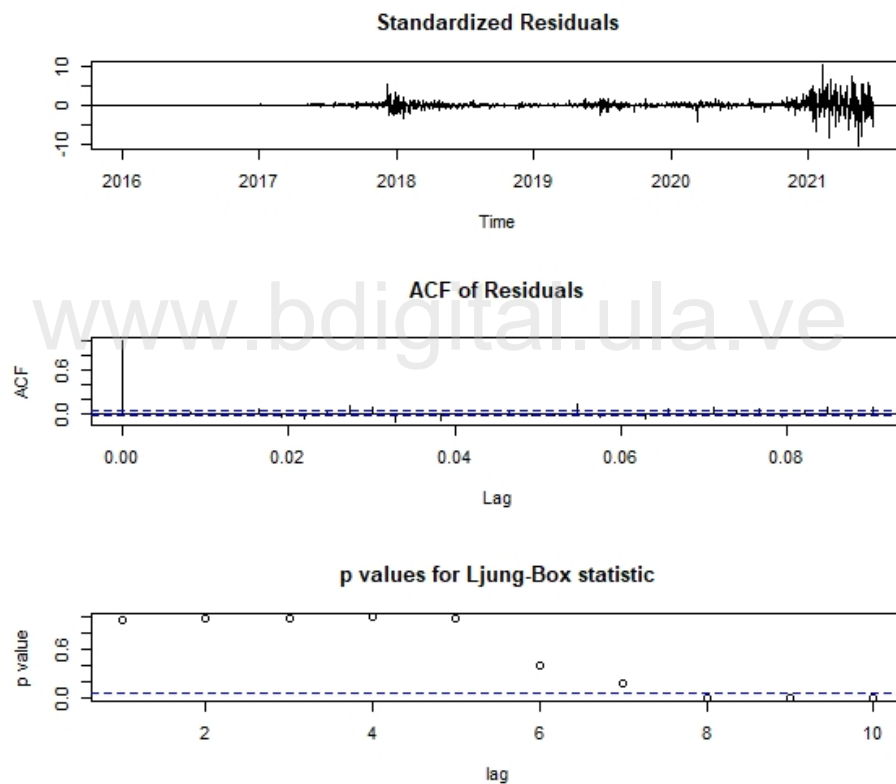


Figura 4.2.16: Gráfica. Residuales del Modelo ARIMA(2,1,1)

Fuente: Elaboración Propia.

En la primera grafica se observa que los residuos estandarizados parecen ser totalmente independientes, ya que se comportan aleatoriamente alrededor del cero.

En la segunda gráfica, se puede ver que la función de autocorrelación simple es un ruido blanco, ya que solo en el *lag* 0 se sale de las bandas de significación y es nula.

En la tercera gráfica se visualiza la prueba de Ljung-Box, visualmente los valores se encuentran fuera del intervalo significativo indicando la no presencia de autocorrelación de los residuales.

Para corroborar lo descrito, se utiliza la prueba de Ljung-Box, que verifica si existe autocorrelación en una serie de tiempo.

H_0 : Los residuos se distribuyen de forma independiente.

H_A : Los residuos no se distribuyen de forma independiente; exhiben correlación serial.

```
> Box.test(residuals(Arima211), type = "Ljung - Box")
Box-Ljung test

data: residuals(Arima211)
X-squared = 0.0010046, df = 1, p-value = 0.9747
```

Figura 4.2.17: Prueba. Ljung Box Test, Residuales del Modelo ARIMA(2,1,1)

Fuente: Elaboración Propia.

El p – *valor* de la prueba es 0.9747, mayor que 0.05. Por lo tanto, no se rechaza la hipótesis nula de la prueba y se concluye que los valores de los datos son independientes, por lo que se puede interpretar que no existen evidencias para rechazar que los residuos sean un ruido blanco.

Pronósticos

Tras validar el modelo ARIMA(2,1,1), se procede a la ejecución de los pronósticos.

```
> prediccion <- forecast(Arima211, h = 30) > plot(prediccion)
```

Nº	Fecha	Precios Reales	Pronósticos	Error
1	23/6/2021	33723,03	32070,89	5%
2	24/6/2021	34662,44	31995,13	8%
3	25/6/2021	31637,78	31912,38	1%
4	26/6/2021	32186,28	31869,93	1%
5	27/6/2021	34649,64	31842,88	9%
6	28/6/2021	34434,34	31826,87	8%
7	29/6/2021	35867,78	31817,17	13%
8	30/6/2021	35040,84	31811,34	10%
9	1/7/2021	33572,12	31807,82	6%
10	2/7/2021	33897,05	31805,70	7%
11	3/7/2021	34668,55	31804,43	9%
12	4/7/2021	35287,78	31803,66	11%
13	5/7/2021	33746,00	31803,20	6%
14	6/7/2021	34235,19	31802,92	8%
15	7/7/2021	33855,33	31802,75	6%
16	8/7/2021	32877,37	31802,65	3%
17	9/7/2021	33798,01	31802,59	6%
18	10/7/2021	33520,52	31802,55	5%
19	11/7/2021	34240,19	31802,53	8%
20	12/7/2021	33155,85	31802,52	4%
21	13/7/2021	32702,03	31802,51	3%
22	14/7/2021	32822,35	31802,50	3%
23	15/7/2021	31780,73	31802,50	0%
24	16/7/2021	31421,54	31802,50	1%
25	17/7/2021	31533,07	31802,50	1%
26	18/7/2021	31796,81	31802,50	0%
27	19/7/2021	30817,83	31802,50	3%
28	20/7/2021	29807,35	31802,50	6%
29	21/7/2021	32110,69	31802,50	1%
30	22/7/2021	32313,11	31802,50	2%

Tabla 4.5: Comparación de los Resultados del Pronósticos de la Serie de Tiempo del Bitcoin

Una vez, generado los pronósticos, se evaluaron y compararon con los precios reales, resultando un error medio de 5% en un periodo de 30 días.

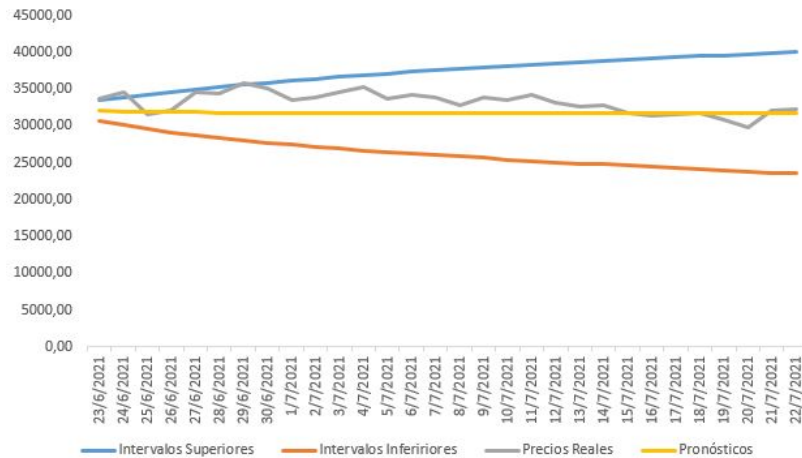


Figura 4.2.17: Gráfica. Pronósticos de la Serie de Tiempo del Bitcoin

Fuente: Elaboración Propia.

La grafica ilustra el comportamiento de los precios reales, pronósticos e intervalos, haciendo enfoques a la precisión del modelo seleccionado.

4.2.2 Ethereum

Se genera la gráfica para visualizar el comportamiento de la serie temporal del Ethereum.

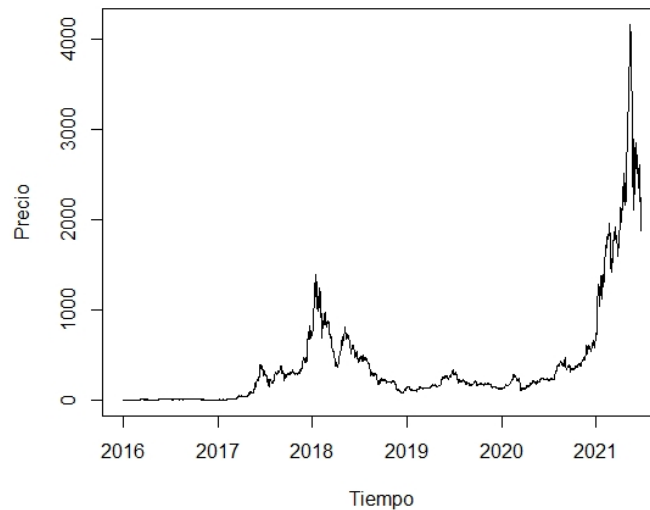


Figura 4.2.18: Gráfica. Serie Original del Ethereum

Fuente: Elaboración Propia.

La grafica presentada muestra el comportamiento del Ethereum, muy similar al comportamiento del Bitcoin, debido que es el principal influyente en el mercado de la criptomonedas. Muestra una tendencia en el primer trimestre del año 2017, conservándose así, hasta llegar a los 1.396,42\$ el 13/01/2018, luego desciende hasta llegar a los 84,31\$ el 14/12/2018, manteniendo una constante fluctuación alrededor de ese valor, no es hasta mediano del año 2020, que se aprecia una tendencia pronunciada llegando a su punto más alto históricamente el 11/05/2021, cotizándose a 4.168,7\$, para luego comenzar a descender como se aprecia en la gráfica.

Claramente la gráfica indica que la serie es no estacionaria, se verifica aplicando la prueba de Dickey-Fuller.

```
Augmented Dickey-Fuller Test
data: Ethe
Dickey-Fuller = -1.6652, Lag order = 12, p-value = 0.7201
alternative hypothesis: stationary
```

Figura 4.2.19: Prueba. Test de Dickey-Fuller, Serie Original del Ethereum

Fuente: Elaboración Propia.

Efectivamente la prueba verifica que la serie es no estacionaria expresando un p -valor = 0.72, mayor que 0.05. Se procede a suavizar la serie aplicando diferenciación a los datos para lograr la estacionariedad de la serie.

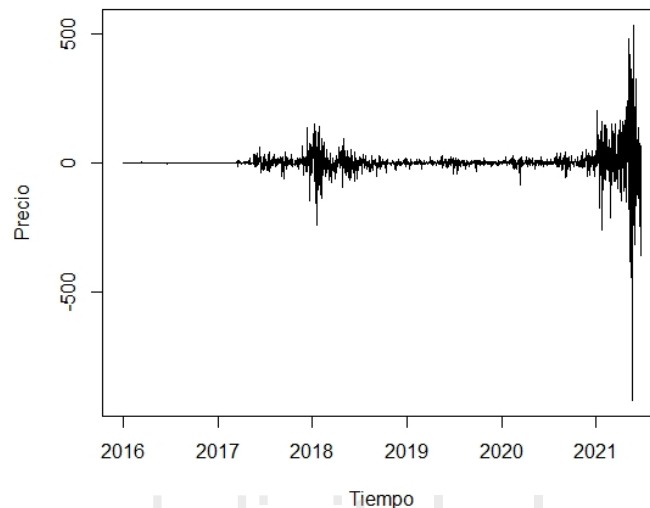


Figura 4.2.20: Gráfica. Serie con Diferenciación de Orden Uno (1) del Ethereum

Fuente: Elaboración Propia.

En la gráfica se observa que la serie diferenciada no muestra tendencia y los valores se encuentran alrededor de cero, es un indicativo que la serie es estacionaria, se utiliza de nuevo la prueba de Dickey-Fuller para verificar la estacionariedad.

```
Augmented Dickey-Fuller Test
data: Ethe1
Dickey-Fuller = -12.784, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary
```

Figura 4.2.21: Prueba. Test de Dickey-Fuller, Serie Diferenciada del Ethereum

Fuente: Elaboración Propia.

Ahora el p -valor = 0.01, rechazando la hipótesis nula, la serie tiene estacionalidad en la primera diferenciación de los términos.

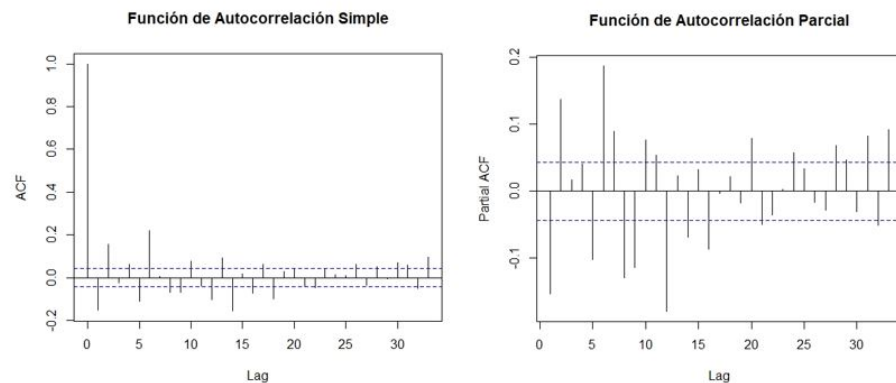


Figura 4.2.22: Gráfica. Funciones ACF y PACF del Ethereum

Fuente: Elaboración Propia.

Se puede observar en la gráfica tanto de la ACF y la PACF, existen varios rezagos que se sobrepasan la banda de significancia, siendo posibles valores para el modelo. Para la elección del modelo al Ethereum, se utiliza la función *auto.arima* de la librería *forecast* de *R*, que proporciona el mejor modelo, en una opción rápida para evaluar entre todos los posibles modelos, considerando diversos criterios como estacionariedad, estacionalidad, diferencias, entre otras.

```
Series: Ethe
ARIMA(5,1,3)

Coefficients:
      ar1      ar2      ar3      ar4      ar5      ma1      ma2      ma3
      -0.8028 -0.7852 -0.2763  0.0709 -0.1279  0.6929  0.9068  0.3257
s.e.    0.0999  0.0671  0.0746  0.0361  0.0329  0.0997  0.0512  0.0860

sigma^2 estimated as 2177: log likelihood=-10514.8
AIC=21047.6  AICc=21047.69  BIC=21098
```

Figura 4.2.23: Resumen. Función *auto.arima*

Fuente: Elaboración Propia.

Tras realizar la función *auto.arima* arrojó, que el mejor modelo para la serie del Ethereum es ARIMA(5,1,3). Se analiza sus residuales para verificar el cumplimiento de ruido blanco.

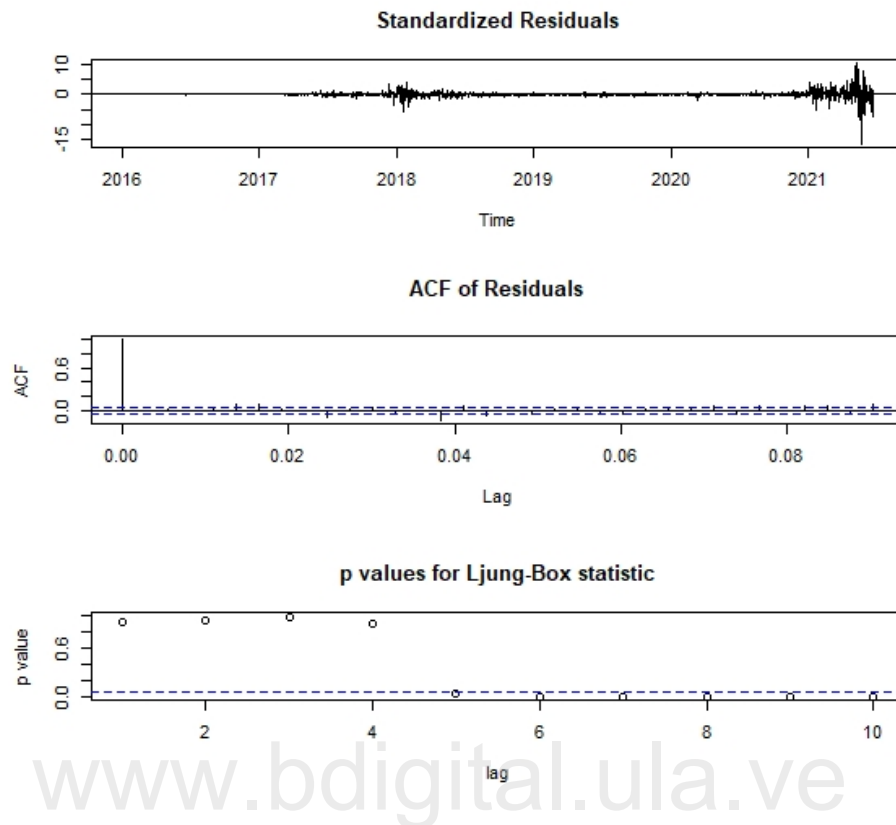


Figura 4.2.24: Gráfica. Residuales del Modelo ARIMA(5,1,3)

Fuente: Elaboración Propia.

Se aprecia en la primera grafica que los residuos estandarizados parecen ser totalmente independientes, ya que se comportan aleatoriamente alrededor del cero.

En la segunda gráfica, se observa que la función de autocorrelación simple es un ruido blanco, ya que solo en el *lag* 0 se sale notoriamente de las bandas de significación y es nula.

En la tercera gráfica se obtiene la prueba de Ljung-Box, visualmente hay valores dentro del intervalo significativo indicando la presencia de autocorrelación de los residuales se verifica a través del test de Ljung-Box.

```
Box-Ljung test  
data: residuals(Arima513)  
X-squared = 0.0078734, df = 1, p-value = 0.9293
```

Figura 4.2.25: Prueba. Ljung Box Test, Residuales del Modelo ARIMA(5,1,3)

Fuente: Elaboración Propia.

El Test Ljung-Box arrojó un $p - valor = 0.9293$, mucho mayor que 0.05, indicando que los residuales se distribuyen de forma independiente, acertando con el ruido blanco

Tras la validación del modelo ARIMA(5,1,3) para el Ethereum, se procede a los pronósticos para un tiempo de 30 días.

www.bdigital.ula.ve

N°	Fecha	Precios Reales	Pronósticos	Error
1	23/6/2021	1989,74	1844,95	8%
2	24/6/2021	1988,46	1843,07	8%
3	25/6/2021	1813,22	1825,33	1%
4	26/6/2021	1829,24	1894,17	3%
5	27/6/2021	1978,89	1852,96	7%
6	28/6/2021	2079,66	1840,59	13%
7	29/6/2021	2160,77	1862,84	16%
8	30/6/2021	2274,55	1873,23	21%
9	1/7/2021	2113,61	1839,11	15%
10	2/7/2021	2150,04	1856,59	16%
11	3/7/2021	2226,11	1869,64	19%
12	4/7/2021	2321,72	1852,75	25%
13	5/7/2021	2198,58	1847,49	19%
14	6/7/2021	2324,68	1866,97	25%
15	7/7/2021	2315,16	1858,82	25%
16	8/7/2021	2120,03	1848,65	15%
17	9/7/2021	2146,69	1859,62	15%
18	10/7/2021	2111,40	1863,11	13%
19	11/7/2021	2139,66	1851,43	16%
20	12/7/2021	2036,72	1855,36	10%
21	13/7/2021	1940,08	1862,49	4%
22	14/7/2021	1994,33	1855,75	7%
23	15/7/2021	1911,18	1853,20	3%
24	16/7/2021	1880,38	1860,34	1%
25	17/7/2021	1898,83	1858,48	2%
26	18/7/2021	1895,55	1853,68	2%
27	19/7/2021	1817,30	1857,70	2%
28	20/7/2021	1787,51	1859,59	4%
29	21/7/2021	1990,97	1855,20	7%
30	22/7/2021	2025,20	1856,03	9%

Tabla 4.6: Comparación de los Resultados del Pronósticos de la Serie de Tiempo del Ethereum

Al obtener los pronósticos se realizó comparaciones con los precios reales, calculando el porcentaje de error en cada uno de esos días como se aprecia en la tabla, con la obtención de estos errores se realizó un error medio, resultando de 11% de error total en ese plazo.

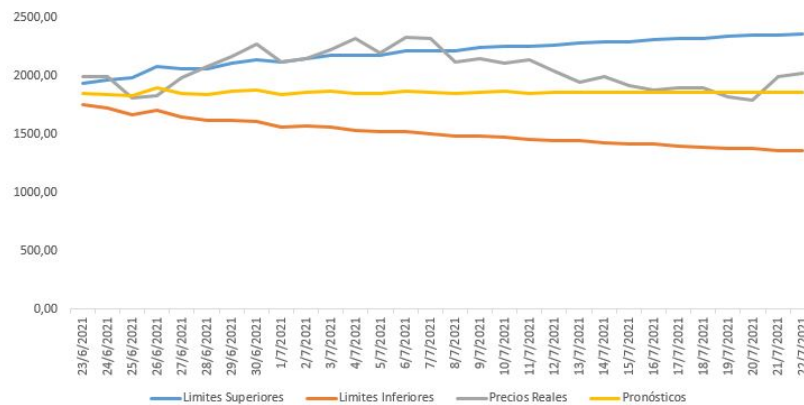


Figura 4.2.26: Gráfica. Pronósticos de la Serie de Tiempo del Ethereum

Fuente: Elaboración Propia.

La gráfica ilustra como los pronósticos y sus intervalos de confianza del modelo $ARIMA(5,1,3)$ para el Ethereum, se mantienen similares en el rango que los precios reales.

4.2.3 Litecoin

El estudio de series temporales del Litecoin, se inicia graficando las observaciones del precio para la descripción del comportamiento.

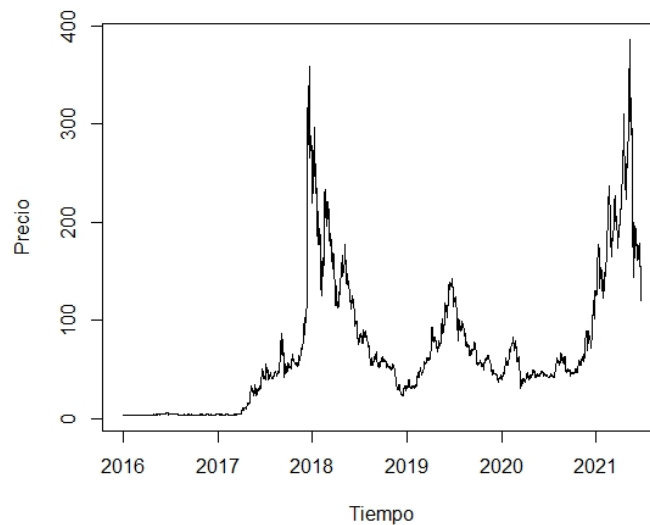


Figura 4.2.27: Gráfica. Serie Original del Litecoin

Fuente: Elaboración Propia.

La grafica muestra un comportamiento similar a las anteriores estudiadas, mostrando una estabilidad en el año 2016, en el primer trimestre del año 2017 presenta un primer incremento hasta posicionarse en 350, 5\$ el 19/12/2017 llegando a uno de sus puntos más altos, comienza un descenso llegando a cotizarse en 23, 78\$ el 15/12/2018, en el año 2019 se mantiene en una fluctuación no tan pronunciada si no es hasta el año 2021 que tiene una tendencia bastante significativa llegando 386, 45\$ el 09/05/2021 el valor más alto de la serie para luego descender, este comportamiento indica que estamos en presencia de una serie no estacionaria, para su verificación se aplica la prueba de Dickey-Fuller.

```
Augmented Dickey-Fuller Test
data: Lit
Dickey-Fuller = -2.8624, Lag order = 12, p-value = 0.2132
alternative hypothesis: stationary
```

Figura 4.2.28: Prueba. Test de Dickey-Fuller, Serie Original del Litecoin

Fuente: Elaboración Propia.

La prueba nos certifica que claramente la serie es no estacionaria con un $p\text{-value} =$

0.21, mayor que 0.05. Se realiza diferenciación a la serie para eliminar variaciones irregulares y optar por estacionariedad.

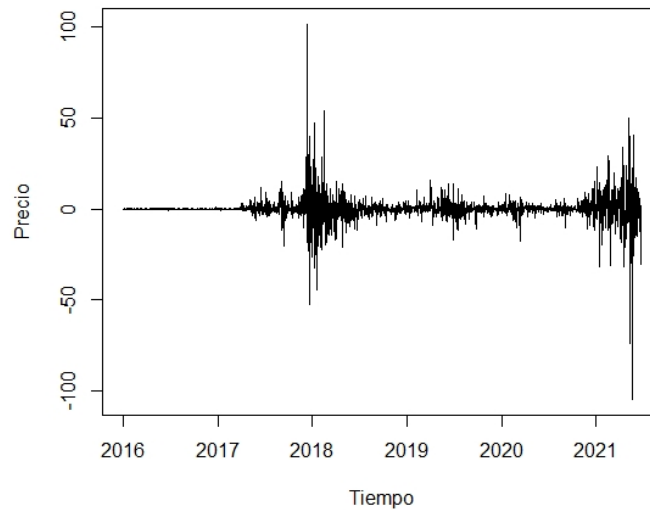


Figura 4.2.29: Gráfica. Serie con Diferenciación de Orden Uno (1) del Litecoin
Fuente: Elaboración Propia.

La grafica no muestra tendencia y los valores se distribuyen alrededor de cero a lo largo del tiempo, se genera de nuevo la prueba de Dickey-Fuller, para verificar si a través de la primera diferenciación se obtiene estacionariedad.

```
Augmented Dickey-Fuller Test
data: Lit1
Dickey-Fuller = -14.108, Lag order = 12, p-value = 0.01
alternative hypothesis: stationary
```

Figura 4.2.30: Prueba. Test de Dickey-Fuller, Serie Diferenciada del Litecoin
Fuente: Elaboración Propia.

La prueba de Dickey-Fuller indico un $p\text{-valor} = 0.01$, rechazando la hipótesis nula, aceptando estacionalidad en la primera diferenciación.

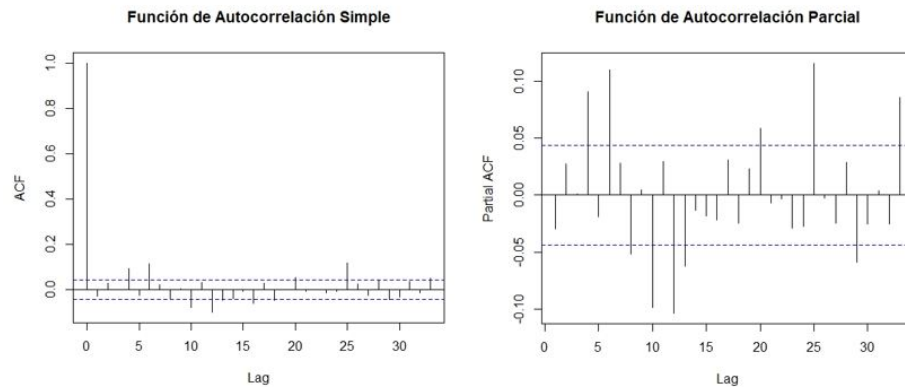


Figura 4.2.31: Gráfica. Funciones ACF y PACF del Litecoin

Fuente: Elaboración Propia.

Las gráficas de función de autocorrelación simple como parcial del Litecoin muestran varios rezagos que sobrepasan las bandas, siendo posibles parámetros para la estimación. Se utiliza la misma función de *auto.arima* siendo una opción rápida para evaluar y elegir entre todos los posibles modelos.

```
Series: Lit
ARIMA(1,1,1)

Coefficients:
      ar1      ma1
    -0.7669  0.7264
s.e.    0.0798  0.0842

sigma^2 estimated as 54.18:  log likelihood=-6825.76
AIC=13657.52  AICC=13657.53  BIC=13674.32
```

Figura 4.2.32: Resumen. Función auto.arima

Fuente: Elaboración Propia.

La función indica que el mejor modelo para el Litecoin es el ARIMA(1,1,1), para validar el modelo se visualiza que los residuales cumplan con un patrón de ruido blanco.

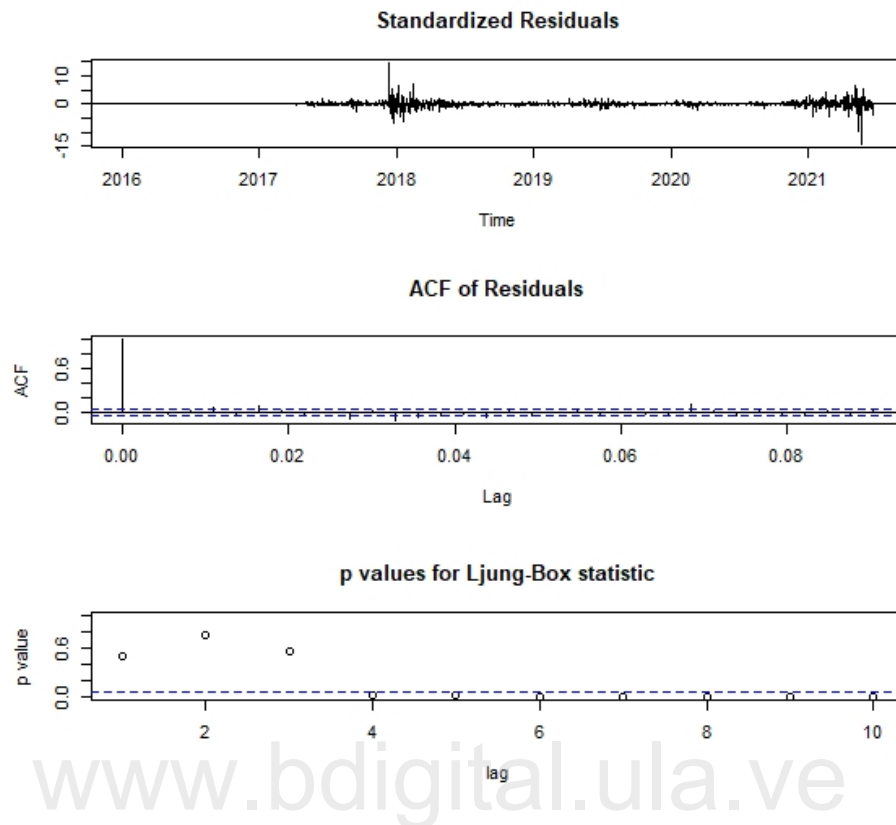


Figura 4.2.33: Gráfica. Residuales del Modelo ARIMA(1,1,1)

Fuente: Elaboración Propia.

En el primer grafico los residuos estandarizados parecen ser independientes, ya que se comportan aleatoriamente alrededor del cero.

En la segunda gráfica, se observa que la función de autocorrelación simple es un ruido blanco, ya que solo en el *lag* 0 se sale notoriamente de las bandas de significación y es nula.

La tercera gráfica se obtiene la prueba de Ljung-Box, visualmente hay valores dentro del intervalo significativo indicando la presencia de autocorrelación de los residuales se verifica a través del test de Ljung-Box.

```
Box-Ljung test  
data: residuals(Arima111)  
X-squared = 0.44145, df = 1, p-value = 0.5064
```

Figura 4.2.34: Prueba. Ljung Box Test, Residuales del Modelo ARIMA(1,1,1)

Fuente: Elaboración Propia.

El Test Ljung-Box arrojó un $p - valor = 0.9293$, mucho mayor que 0.05, acertando con el ruido blanco, los residuales se distribuyen de forma independiente. Una vez, validado el modelo se procede a los pronósticos.

www.bdigital.ula.ve

N°	Fecha	Precios Reales	Pronósticos	Error
1	23/6/2021	128,67	119,07	8%
2	24/6/2021	134,48	119,48	13%
3	25/6/2021	125,69	119,17	5%
4	26/6/2021	126,66	119,40	6%
5	27/6/2021	132,35	119,22	11%
6	28/6/2021	137,37	119,36	15%
7	29/6/2021	143,7	119,25	20%
8	30/6/2021	144,14	119,34	21%
9	1/7/2021	137,58	119,27	15%
10	2/7/2021	136,94	119,32	15%
11	3/7/2021	140,28	119,28	18%
12	4/7/2021	144,91	119,31	21%
13	5/7/2021	138,07	119,29	16%
14	6/7/2021	138,99	119,31	16%
15	7/7/2021	137,37	119,29	15%
16	8/7/2021	132,31	119,30	11%
17	9/7/2021	134,57	119,30	13%
18	10/7/2021	133,99	119,30	12%
19	11/7/2021	134,25	119,30	13%
20	12/7/2021	133,54	119,30	12%
21	13/7/2021	131,32	119,30	10%
22	14/7/2021	131,06	119,30	10%
23	15/7/2021	125,49	119,30	5%
24	16/7/2021	121,04	119,30	1%
25	17/7/2021	120,03	119,30	1%
26	18/7/2021	119,51	119,30	0%
27	19/7/2021	113,16	119,30	5%
28	20/7/2021	107,4	119,30	10%
29	21/7/2021	117,46	119,30	2%
30	22/7/2021	120,74	119,30	1%

Tabla 4.7: Comparación de los Resultados del Pronósticos de la Serie de Tiempo del Litecoin

Se realizó comparaciones de los pronósticos del modelo ARIMA(1,1,1) y los precios reales del Litecoin, calculando el porcentaje de error en cada uno de los días, resultando con un error medio total de de 11%.

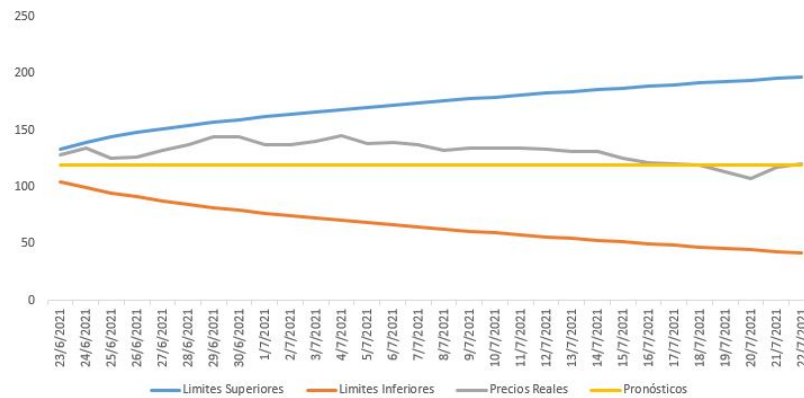


Figura 4.3.14: Gráfica. Pronósticos de la Serie de Tiempo del Litecoin

Fuente: Elaboración Propia.

La grafica describe visualmente el ajuste del modelo ARIMA(1,1,1) para el Litecoin.

Capítulo 5

Conclusiones y Recomendaciones

Partiendo que muchos analistas consideran las criptomonedas uno de los activos más rentables para invertir, se crea la necesidad de optar por pronósticos eficaces que permitan confiabilidad a la hora de seleccionar una criptomoneda como inversión.

Como se ha observado en el documento, la selección de los modelos estadísticos para predecir las criptomonedas fueron la regresión lineal simple y series de tiempo, que otorgaron de manera eficaz pronósticos con resultados fiables de la evolución futura de 30 días a las criptomonedas seleccionadas.

Para los pronósticos de regresión lineal simple se utilizó el método de mínimos cuadrados, que permite mediante un conjunto de datos determinar la función continua que mejor se aproxime a los datos (línea de regresión o la línea de mejor ajuste). Resultando para la criptomoneda del Bitcoin un error medio total de 35%, para el Ethereum de 113% y 6% para el Litecoin, en un lapso de 30 días, estas diferenciaciones de cada una de las criptomonedas seleccionadas se deben a la dispersión de los datos. Sin embargo, la regresión lineal como modelo predictivo para las criptomonedas seleccionadas es bastante efectiva para el Litecoin, resultando un porcentaje de error bastante bajo, considerando las altas fluctuaciones que presentan estos critoactivos.

El pronóstico de serie de tiempo para los precios de las criptomonedas fue una tarea

difícil y compleja por estar presente diferentes niveles de estacionalidad y la presencia de variables exógenas. El precio de las criptomonedas es una variable afectada por otras de tipo cuantitativo y cualitativo, por esta razón, se consideró utilizar a los modelos univariantes llamados ARIMA; es decir, que los propios datos temporales históricos de la variable estudiada indican la estructura probabilística y el patrón teórico de comportamiento. Los modelos ARIMA resultaron altamente efectivos, los pronósticos más acertado fue para el Bitcoin con un error medio total de 5%, con un 95% de efectividad, sin embargo, el Ethereum y Litecoin resultaron con un error medio total de 11%, siendo los modelos ARIMA más efectivos que la regresión lineal.

En cuanto a la investigación sobre la eficacia de los modelos de predicción, muestra que es posible predecir con altos niveles de exactitud el comportamiento futuro del precio de las criptomonedas mediante la implementación de estos modelos, facilitando así la toma de decisiones de inversión en las distintas criptomonedas.

Los análisis realizados en este proyecto, son recomendables para todo aquel inversionista, comerciante o minador de criptomonedas, dado que con la ayuda de esta herramienta podrían determinar cuál es el mejor momento para comprar o vender sus criptoactivos.

Bibliografía

- [1] Acerca de ripple. <https://btcdirect.eu/es-es/acerca-de-ripple>. Accedido 27-06-2019.
- [2] Bitcoin. <https://www.criptonoticias.com/criptopedia/blockchains-criptomonedas-fundamentos-caracteristicas/#Bitcoin>. Accedido 11-03-2019.
- [3] Bitcoin un sistema de efectivo electrónico usuario a usuario. https://bitcoin.org/files/bitcoin-paper/bitcoin_es_latam.pdf. Accedido 23-03-2018.
- [4] Coinmarketcap. <https://coinmarketcap.com/>. Accedido 21-02-2019.
- [5] Cripto fácil. <https://www.criptofacil.lat/cripto-noticias/monedas/191-cuales-son-las-ventajas-y-desventajas-de-las-criptomonedas.html>. Accedido 23-02-2019.
- [6] Cripto minería. <https://cripto-mineria.com/criptomonedas/eos#n-2>. Accedido 01-07-2019.
- [7] Criptonoticias. <https://www.criptonoticias.com/opinion/blockchain-la-revolucion-de-la-descentralizacion/>. Accedido 29-02-2019.
- [8] Integridad de datos gracias a blockchain. <https://www.academiablockchain.com/2018/08/07/integridad-de-datos-gracias-a-blockchain/>. Accedido 30-06-2019.

- [9] La regulación de las criptomonedas. <https://criptomonedastop.com/la-regulacion-las-criptomonedas/>. Accedido 04-06-2019.
- [10] La viabilidad de las criptomonedas. <http://www.eldiariodecoahuila.com.mx/tecnologia/2018/2/23/analiza-experto-cinvestav-viabilidad-criptomonedas-714604.html>. Accedido 22-05-2019.
- [11] Litecoin: La plata de las criptomonedas. <https://www.criptonoticias.com/colecciones/litecoin-plata-criptomonedas/>. Accedido 30-06-2019.
- [12] Qué es el análisis técnico. <https://masterthecrypto.com/fundamental-analysis-for-cryptocurrencies/?lang=es>. Accedido 14-02-2019.
- [13] Qué es ethereum eth. <https://www.criptonoticias.com/criptopedia/que-es-ethereum-eth/>. Accedido 28-06-2019.
- [14] Qué es la tecnología de contabilidad distribuida o blockchain. <https://www.criptonoticias.com/criptopedia/que-es-tecnologia-contabilidad-distribuida-blockchain/>. Accedido 22-06-2019.
- [15] Qué es un wallet o monedero digital de bitcoin y criptomonedas. <https://guiabitcoin.com/monederos>. Accedido 26-04-2019.
- [16] Quiénes están detrás del desarrollo de las criptomonedas. <https://www.criptonoticias.com/criptopedia/quienes-desarrollo-bitcoin-criptomonedas/>. Accedido 28-06-2019.
- [17] Son las criptomonedas el nuevo dinero fiduciario. <https://efxto.com/son-las-criptomonedas-el-nuevo-dinero-fiduciario>. Accedido 25-06-2019.
- [18] Tecnología informática. <https://tecnologia-informatica.com/>. Accedido 20-03-2019.

- [19] Una guía para el análisis fundamental para cryptocurrencies. <https://masterthecrypto.com/fundamental-analysis-for-cryptocurrencies/?lang=es>. Accedido 20-02-2019.
- [20] Federico González. Análisis predictivo en bitcoin utilizando técnicas de aprendizaje profundo,. <https://hdl.handle.net/20.500.12008/20471>, 2019. Accedido 28-06-2019.
- [21] Luca Maria Aiello y Andrea Baronchelli Laura Alessandretti, Abeer ElBahrawy. Anticipating cryptocurrency prices using machine learning. <https://www.hindawi.com/journals/complexity/2018/8983590/>, 2018. Accedido 17-01-2019.
- [22] Raúl Andrés Marqués López. Criptomonedas, ¿evolución o disrupción financiera? : una propuesta de previsión basado en las cadenas de markov = cryptocurrencies, evolution or financial disruption? : a prevision proposal based on markov chains. <https://buleria.unileon.es/handle/10612/8566>, 2018. Accedido 04-02-2019.