



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



**“ELECCIÓN DEL MEJOR MODELO ENTRE REGRESIÓN
LINEAL MÚLTIPLE Y ÁRBOLES DE REGRESIÓN PARA
PREDECIR EL PRECIO MÁXIMO DE LAS ACCIONES DE INTEL
EN FUNCIÓN AL PRECIO DE APERTURA Y VOLUMEN DE
VENTAS DE ACCIONES POR DIA - 2019.”**

TESIS

PRESENTADA POR:

Bach. ALICIA ROXANA MAYDANA HUANCA

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

PUNO – PERÚ

2021



DEDICATORIA

A Dios por darme la vida y su infinita bondad, por darme sabiduría e inteligencia, quien supo guiarme por el buen camino y por darme fuerzas para seguir adelante.

Con cariño a mis padres, Emeterio Maydana y Cipriana Huanca, por el apoyo incondicional, por ser la base fundamental en mi formación personal y profesional, por enseñarme a crecer ante las adversidades y afrontar los desafíos día a día, de la misma forma a cumplir mis y objetivos trazados.

A mis hermanos Edgar, Sandra Indira, Ramiro y Elmer, por el apoyo moral, fueron quienes me motivaron e influyeron en mí, demostrándome la superación profesional y la perseverancia.

ALICIA MAYDANA



AGRADECIMIENTOS

- A la Universidad Nacional del Altiplano de Puno, en especial a la Facultad de Ingeniería Estadística e Informática, a los docentes por haber compartido sus sabios conocimientos a lo largo de mi formación profesional.
- A los miembros de jurado al Dr. Juan Reynaldo Paredes Quispe, a la Dr. Maria Maura Salas Pilco y al M. Sc. Leonid Aleman Gonzales, por sus contribuciones para la realización de esta investigación.
- Al Dr. Edgar Eloy Carpio Vargas, asesor de tesis y sobre todo una gran persona; por la paciencia, orientación y guía en el desarrollo de esta investigación.
- A mis amigos, por hacer de esta etapa una de las más bonitas que he vivido y convertirse en personas esenciales en mi vida, a pesar de la distancia, me han transmitido constantemente palabras de aliento y apoyo moral.
- Agradezco a las personas, quienes acotaron en el desarrollo de este Trabajo de Investigación, que sin su ayuda no hubiera sido posible este logro.



ÍNDICE GENERAL

Pág.

DEDICATORIA

AGRADECIMIENTOS

ÍNDICE GENERAL

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

ÍNDICE DE ACRÓNIMOS

RESUMEN 10

ABSTRACT..... 11

CAPÍTULO I

INTRODUCCIÓN

1.1. DESCRIPCIÓN DEL PROBLEMA 14

1.2. FORMULACIÓN DEL PROBLEMA 15

1.3. JUSTIFICACIÓN DE LA INVESTIGACIÓN 15

1.4. OBJETIVOS DE LA INVESTIGACIÓN 16

1.4.1. Objetivo General 16

1.4.2. Objetivos Específicos 16

1.5. HIPOTESIS DE LA INVESTIGACIÓN 16

1.5.1. Hipótesis General 16

CAPÍTULO II

REVISIÓN DE LITERATURA

2.1. ANTECEDENTES 17

2.1.1. Nivel Internacional 17

2.1.2. Nivel Nacional 18

2.1.3. Nivel Regional 19



2.2. MARCO TEÓRICO	19
2.2.1. Regresión lineal múltiple	19
2.2.2. Árboles de regresión	36
2.3. MARCO CONCEPTUAL	41
2.3.1. Volumen.....	41
2.3.2. Precio Máximo.....	41
2.3.3. Precio de Apertura	41
2.3.4. Acciones Ordinarias.....	41

CAPÍTULO III

MATERIALES Y MÉTODOS

3.1. TIPO Y DISEÑO DE INVESTIGACIÓN	42
3.1.1. Tipo de investigación.....	42
3.1.2. Diseño de investigación	42
3.2. UBICACIÓN DEL TRABAJO DE INVESTIGACIÓN.....	42
3.3. POBLACIÓN Y MUESTRA.....	42
3.3.1. Población	42
3.3.2. Muestra	42
3.4. ALCANCE DE LA INVESTIGACIÓN	43
3.5. TÉCNICA E INSTRUMENTOS DE RECOLECCIÓN DE DATO	44
3.5.1. Técnica.....	44
3.5.2. Instrumento:	44
3.6. METODOLOGÍA ESTADÍSTICA UTILIZADA (procedimiento)	44
3.7. ELABORACIÓN DE REGRESIÓN LINEAL MÚLTIPLE	45
3.8. ELABORACIÓN DEL ARBOL DE REGRESIÓN.....	45
3.9. COMPARACIÓN DE ERRORES.....	46



CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

3.1. ANÁLISIS DESCRIPTIVO	47
3.2. MODELO DE REGRESIÓN LINEAL MÚLTIPLE	52
3.2.1. Obtención del primer modelo de RLM.....	52
3.2.2. Obtención del segundo modelo de RLM.	59
3.2.3. Obtención del tercer modelo de RLM.	62
3.3. OBTENCIÓN DEL ARBOL DE REGRESIÓN.	67
3.3.1. Árbol regresión utilizando el tipo uniforme:	68
3.3.2. Árbol de regresión algoritmo uniforme (<i>Poda</i>)	71
3.4. RESUMEN DE LOS ALGORITMOS	72
V. CONCLUSIONES.....	74
VI. RECOMENDACIONES	75
VII. REFERENCIAS BIBLIOGRÁFICAS.....	76
ANEXOS.....	80

Área : Análisis multivariado, big data, minería de datos e investigación de mercados.

Tema : Regresión multivariada.

FECHA DE SUSTENTACIÓN: 03 de marzo 2021



ÍNDICE DE TABLAS

	Pág.
Tabla 1. Estadísticos descriptivos de las variables en estudio.	47
Tabla 2. Correlaciones de las variables Precio máximo, Precio de apertura y Volumen a partir de Intel 2019.	50
Tabla 3 . Resumen de los supuestos del primer modelo de regresión	58
Tabla 5. Resumen de los supuestos del tercer modelo de regresión.....	66
Tabla 6. Resumen de errores de la Regresión Lineal Múltiple.....	67
Tabla 7. Comparación de errores del Árbol de regresión	72
Tabla 8. Resumen de errores para la Regresión lineal y Árbol de regresión.....	72



ÍNDICE DE FIGURAS

	Pág.
Figura 1. Regresión lineal y su interpretación geométrica	20
Figura 2. Prueba de contraste de la hipótesis "t"	32
Figura 3. Elementos del Árbol de regresión	38
Figura 4. Particiones del algoritmo CART	39
Figura 5. Ilustración gráfica de la regresión lineal y el árbol de decisión	40
Figura 6. Distribución de la variable precio máximo	48
Figura 7. Distribución en la variable precio de apertura.	48
Figura 8. Distribución en la variable volumen	49
Figura 9. Gráfico de cajas de la variable precio máximo de las acciones de Intel 201950	
Figura 10. Gráfico de cajas de la variable Precio Máximo de las Acciones de Intel 2019	51
Figura 11. Gráfico de cajas de la variable Volumen de ventas de las acciones de Intel.	52
Figura 12. Linealidad de las variables precio apertura y volumen de las acciones Intel	54
Figura 13. Gráfico de cuantiles para los errores del modelo	55
Figura 14. Gráfico de identificación de valores atípicos	58
Figura 15. Gráfico de datos Estandarizados mayores a 1.5	61
Figura 16. Gráfico de linealidad de las variables precio de apertura y volumen.....	63
Figura 17. Numero de hojas entrenamiento y validación	67
Figura 18. Árbol de Regresión utilizando el tipo Uniforme para predecir.....	68
Figura 19. Representación del Árbol de Regresión podado con el algoritmo uniforme.	71



ÍNDICE DE ACRÓNIMOS

AR:	Árbol de Regresión
CART:	Classification and Regression Trees
MCO:	Mínimos Cuadrados Ordinarios
MELI:	Mejor Estimador Lineal Insesgado
MSE:	Error Cuadrático Medio
NASDAQ:	National Association of Securities Dealers Automated Quotations.
RLM:	Regresión Lineales Múltiple
VIF:	Factor de Inflación de la Varianza



RESUMEN

La Investigación tuvo como objetivo comparar el Modelo de regresión lineal Múltiple frente al Árbol de regresión, para ello se utilizó las variables Precio máximo de las acciones de Intel en función al Precio de apertura y Volumen de ventas, de acciones por día. El diseño de investigación fue no experimental y de corte transversal el tipo de investigación descriptivo correlacional, en la cual la población estuvo conformada por todas las acciones de la empresa Intel desde su creación y a través del tiempo; se empleó el muestreo no probabilístico por conveniencia, se consideró desde mayo del 2018 hasta octubre del 2019 siendo un total de 410 registros recopilados a partir de la revisión documentaria. Las pruebas estadísticas usadas fueron el Análisis de regresión lineal múltiple y los Árboles de regresión. Los resultados obtenidos fueron; el Modelo de Regresión lineal múltiple con la técnica de eliminación de datos atípicos queda definida por la siguiente ecuación $Y = 0.02856 + 1.003X_1 + 0.000000009405X_2$. Alcanzando una prueba F significativa y la bondad de ajuste es bastante alta $R^2 = 0.9979$, y un Error Estándar Residual de 0.2257 dólares, El Árbol de regresión establece que la variable para explicar el Precio máximo de acciones es el Precio de apertura, eliminando la variable volumen, el Error Medio Cuadrático es de 1.4480 dólares. Finalmente se concluye que el mejor modelo para predecir el precio máximo de acciones de Intel es el modelo de Regresión Lineal Múltiple con eliminación de puntos Outliers,

Palabras clave: Regresión, árboles de decisión, precio de acciones.



ABSTRACT

The objective of the Investigation was to compare the Multiple linear regression Model against the Regression Tree, for this, the variables Maximum price of Intel shares based on the Opening Price and Sales Volume, of shares per day were used. The research design was non-experimental and cross-sectional, the type of correlational descriptive research, in which the population was made up of all the actions of the Intel company since its creation and over time; Non-probability sampling for convenience was used, it was considered from May 2018 to October 2019, a total of 410 records collected from the documentary review. The statistical tests used were Multiple Linear Regression Analysis and Regression Trees. The results obtained were; The Multiple Linear Regression Model with the Outliers elimination technique is defined by the following equation $Y = 0.02856 + 1.003X_1 + 0.000000009405X_2$. Reaching a significant F-test and the goodness of fit is quite high $R^2 = 0.9979$, and a Residual Standard Error of \$ 0.2257, The Regression Tree establishes that the variable to explain the Maximum Stock Price is the Opening Price, eliminating the variable volume, the root mean square error is \$ 1.4480. Finally, it is concluded that the best model to predict the maximum price of Intel shares is the Multiple Linear Regression model with elimination of Outliers points,

Keywords: Regression, decision trees, stock price.



CAPÍTULO I

INTRODUCCIÓN

La toma de decisión es esencial en toda organización o empresa, por ello es imprescindible buscar técnicas y metodologías estadísticas, que sean precisas y que ayuden a resolver problemas de incertidumbre, estas surgen a la hora de tomar decisiones.

Cada día se generan grandes cantidades de datos que necesitan ser tratados mediante Metodologías, que generan información, que nos pueden ayudar a Investigar, Predecir o tomar decisiones, además el desarrollo tecnológico nos permite no solo almacenar la Información sino procesarlo y generar conocimiento. (Lizares, 2017).

Para dar apoyo en el ámbito empresarial; en beneficio de la Organización, Empresa o corporación, sobre todo para los inversionistas, accionistas u otros interesados. Por ello se pretende optar por una técnica estadística, la que mejor aprenda el comportamiento de los datos, los mismos que serán tomados básicamente de la información documental.

La regresión lineal múltiple suele ser muy tajante con los supuestos teóricos, por ello, se pretende construir un modelo matemático que ayude al inversor o accionista a poder vender su bien, constantemente se busca alternativas que sean lo suficientemente robustas, en este caso el árbol de regresión puede ser una excelente alternativa, muchas veces logra captar el comportamiento de los datos, incluso no necesariamente está siendo Lineal.

El trabajo de investigación pretende comparar la Regresión Lineal Múltiple frente al Árbol de Regresión, para predecir el Precio Máximo que alcanza una Acción, en función al Precio en el cual es abierto al mercado y el Volumen de ventas de las Acciones en un determinado tiempo, en este sentido el estudio se realiza a la empresa Intel.

La investigación se tiende a desarrollar bajo la siguiente estructura:



Capítulo I: Se Presenta la descripción del problema, formulación del problema, Justificación de esta, como también se muestran los Objetivos y la Hipótesis de dicha Investigación.

Capítulo II: Se detalla el Marco Teórico, fundamentos y/o bases teóricas de la Investigación, Conceptos Básicos y Antecedentes referentes al tema de Investigación Elección del mejor Modelo entre Regresión Lineal Múltiple y Árboles de Regresión.

Capítulo III: Se presenta Materiales y Métodos, donde se detalla la Población y Muestra, Diseño, Alcance de la Investigación, Técnicas e Instrumentos de Recolección de Datos, Metodología Estadística utilizada.

Capítulo IV: Se exponen los Resultados y la Discusión de estos, constituido de cada objetivo planteado en la Investigación, demostrando los resultados mediante cuadros y gráficos los mismos que han sido analizados, para poder validar las Hipótesis Planteadas al inicio, de esta forma llega a las Conclusiones y Recomendaciones.

Finalmente se presentan, las referencias bibliográficas, anexos y como también se Muestra la Base de Datos.



1.1. DESCRIPCIÓN DEL PROBLEMA

Las acciones de las empresas, son un negocio en muchos empresarios y accionistas, las mismas que varían constantemente dependiendo de cuan reconocida sea cada empresa o corporación.

Se han desarrollado diversos métodos para modelar variables de comportamiento continuo frente a la Regresión lineal múltiple; cuyos resultados dependen de que tan bien los datos se ajusten a los supuestos teóricos, que estas mismas tienden a presentar. Los árboles de regresión son una excelente alternativa para ello, se utilizó el Precio de apertura y el volumen de ventas de la empresa Intel para predecir el Precio máximo que este alcance en un determinado periodo de tiempo, y medir en términos de error si el árbol de regresión es capaz de enfrentar este tipo de datos y compararlo frente a la Regresión lineal múltiple.

La empresa Intel se dedica a la venta y fabricación de circuitos integrados, además es el mayor fabricante de estos. Sin embargo, para los accionistas e inversionistas de dicha empresa es muy riesgoso vender acciones sin saber si esta le generará ganancias o pérdidas de dinero al vender su bien, en este sentido, se plantea predecir el Precio máximo que una acción puede tomar durante un determinado tiempo, de esta forma los accionistas e inversionistas tendrán una mejor aproximación, sobre cuanto sería el Precio máximo que alcanzaría una acción.

De esta manera se reducirá las pérdidas de capital y la incertidumbre en accionistas de la empresa Intel, al realizar una venta de su bien.



1.2. FORMULACIÓN DEL PROBLEMA

Para ello se plantea la siguiente interrogante:

¿Tendrá un mejor desempeño la Regresión lineal múltiple frente al Árbol de regresión para predecir el precio máximo de las acciones de Intel en un determinado lapso de tiempo utilizando el precio al cual se apertura al mercado y el volumen de ventas en un determinado periodo de tiempo?

1.3. JUSTIFICACIÓN DE LA INVESTIGACIÓN

El trabajo de Investigación se efectuó para evaluar el modelo de Regresión lineal múltiple frente a un Árbol de regresión, que realice una mejor predicción de los precios de las Acciones de la Empresa Intel.

La regresión lineal múltiple trata de ajustar modelos lineales o linealizables entre una variable dependiente y más de una variable independiente y el árbol de regresión permite escoger convenientemente, una estrategia cuando se enfrentan a diversas alternativas y eventos inciertos al momento de tomar una decisión.

Si bien hoy en día, existe una gran diversidad de técnicas y metodologías estadísticas para predecir el costo de las acciones de las empresas en el mercado, se decidió realizar este estudio porque no existen estudios suficientes que comparen el Modelo de regresión lineal múltiple frente al Árbol de regresión, para predecir el precio de las acciones bursátiles en este caso de la empresa Intel.

Los resultados de este trabajo de investigación servirán a los inversionistas y accionistas para tomar decisiones acertadas, del mismo modo maximizaran ganancias, también servirá como antecedente a futuros trabajos de Investigación relacionados al tema en cuestión.



1.4. OBJETIVOS DE LA INVESTIGACIÓN

1.4.1. Objetivo General

Determinar el mejor modelo entre Regresión lineal múltiple y Árboles de regresión para predecir el precio máximo de las acciones de Intel en función al precio de apertura y el volumen de ventas de acciones, por día en el 2019.

1.4.2. Objetivos Específicos

- Obtener un modelo de Regresión lineal múltiple para predecir el precio máximo del valor de las acciones en función al precio de apertura y el volumen de ventas de las acciones de Intel en el 2019.
- Obtener un Árbol de regresión para predecir el precio máximo de las acciones en función al precio de apertura y el volumen de ventas de las acciones de Intel en el 2019.
- Comparar el modelo de Regresión lineal múltiple y el Árbol de regresión para predecir el precio máximo de las acciones en función al precio de apertura y el volumen de ventas de las acciones de Intel en el 2019.

1.5. HIPOTESIS DE LA INVESTIGACIÓN

1.5.1. Hipótesis General

El modelo de Regresión lineal múltiple tiene menor error frente a un Árbol de Regresión para predecir las Acciones de Intel en el 2019.



CAPÍTULO II

REVISIÓN DE LITERATURA

2.1. ANTECEDENTES

2.1.1. Nivel Internacional

Dupouy (2014) halló que la inclusión de estas técnicas en el campo de las finanzas si tiene una historia reciente y son muchas sus posibles aplicaciones dada las características específicas de las series financieras en términos de volumen, frecuencia, facilidad de obtención y calidad haciéndolas candidatas naturales para su uso en minería de datos para obtener conocimientos de su interacción.

Espinosa (2015) demostró que, realizando un análisis técnico de los datos históricos del valor de las acciones durante un plazo determinado, es posible predecir el cambio de precio de una acción, con la finalidad de ayudar al inversor al momento de tomar la decisión de compra-venta trading de sus acciones, utilizando para esto regresión múltiple.

Peña (2011) indicó que para el grupo de empresas seleccionadas en su investigación en el periodo 2002-2008, a través del modelo CAPM, se encontró que la rentabilidad de las empresas pertenecientes al sector financiero son las que tienen mayor asociación con comportamiento del mercado, es decir que los cambios en el mercado accionario a través del Índice General de la Bolsa de Valores de Colombia se ven reflejados en el rendimiento de las empresas que pertenecen a ese sector y que fueron tomadas en este estudio.

García (2010) concluyó que, las investigaciones anteriores aplicadas en diferentes mercados muestran que existen anomalías relacionadas con efectos calendario. Sin embargo, para el caso colombiano los trabajos se encuentran enfocados



principalmente al estudio de los índices locales. Así, este trabajo pretende dar una visión particular sobre el efecto día en la acción de Ecopetrol, uno de los activos financieros más importantes del mercado de valores colombiano.

Rizola (2017) halló que las caídas bruscas de los índices y los eventos anormales se presentan con inusitada frecuencia en los mercados bursátiles. El mercado mexicano, representado por su principal índice, el IPC, no puede ser la excepción. Se demuestra que en ciertos periodos se han llegado a registrar significativos cambios en los rendimientos que rebasan fácilmente las tres sigmas llegando a alcanzar, en 2008, más de siete. La probabilidad de que se presenten esos hechos es realmente pequeña, prácticamente imposible.

2.1.2. Nivel Nacional

Leon (2018) indicó que las técnicas para el desbalanceo de clases el cual ha permitido obtener dos modelos para los casos de fracaso. Para ambos casos, se han encontrado buenos resultados aplicando el algoritmo Random Forest. Finalmente, por el método AUC se permitió identificar al mejor clasificador, por lo que se escogió Random Forest.

Giménez & Zamorano (2014) indicaron que hacer pronósticos precisos de índices bursátiles es prácticamente imposible. Incluso teniendo modelos que parecieran dar buenos resultados, como los que elegimos finalmente, éstos pueden presentar problemas en la realidad debido al sobreajuste o “overfitting”.

Alvarez, (2016) indicó que la metodología Ítem a Ítem planteada por Amazon resulto útil para predecir el orden de preferencia de los ligandos que se unen a un determinado blanco sin embargo los resultados predictivos no resultaron suficientemente buenos. Si bien la medida aprendió la tendencia para cada proteína esta no pudo generalizar la información.

Camborda (2014) determinó que según el algoritmo generado y de este grupo de datos donde se tomó en cuenta el mejor atributo que tenían mayor ganancia de información en cada grupo de variables con tan solo dos tipos de salidas, la variable que predomina es una variable de tipo académica que es como ha sido el rendimiento académico en la secundaria.

2.1.3. Nivel Regional

Candia (2019) comprobó que según los resultados encontrados, el algoritmo de árboles de decisión “Random Forest”, fue el algoritmo que tuvo el mejor performance para la predicción del rendimiento académico de los ingresantes en los primeros semestres a la UNSAAC con un 69% de predicción, el segundo algoritmo con mejor performance fue algoritmo de Regresión Logística con un 68% para el presente caso de estudio.

2.2. MARCO TEÓRICO

2.2.1. Regresión lineal múltiple

La regresión lineal múltiple permite generar un modelo lineal en el que el valor de la variable dependiente o respuesta Y se determina a partir de un conjunto de variables independientes llamadas predictores X 's. Es una extensión de la regresión lineal simple, por lo que es fundamental comprender esta última. Los modelos de regresión múltiple pueden emplearse para predecir el valor de la variable dependiente o para evaluar la influencia que tienen los predictores sobre ella (esto último se debe que analizar con cautela para no malinterpretar causa-efecto).

Los modelos lineales múltiples siguen la siguiente ecuación:

$$Y_i = (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_{ni}) + e_i \quad i = 1, 2, 3, \dots, n \quad (1)$$

Donde:

Y_i : Es la variable dependiente o respuesta.

β_0 : Es la ordenada en el origen, el valor de la variable dependiente Y cuando todos predictores son cero.

β_i : Es el efecto promedio que tiene el incremento en una unidad de la variable predictora

X_i : Es la variable independiente o explicativas

ϵ_i : Es el residuo o error, la diferencia entre el valor observado y el estimado por el modelo.

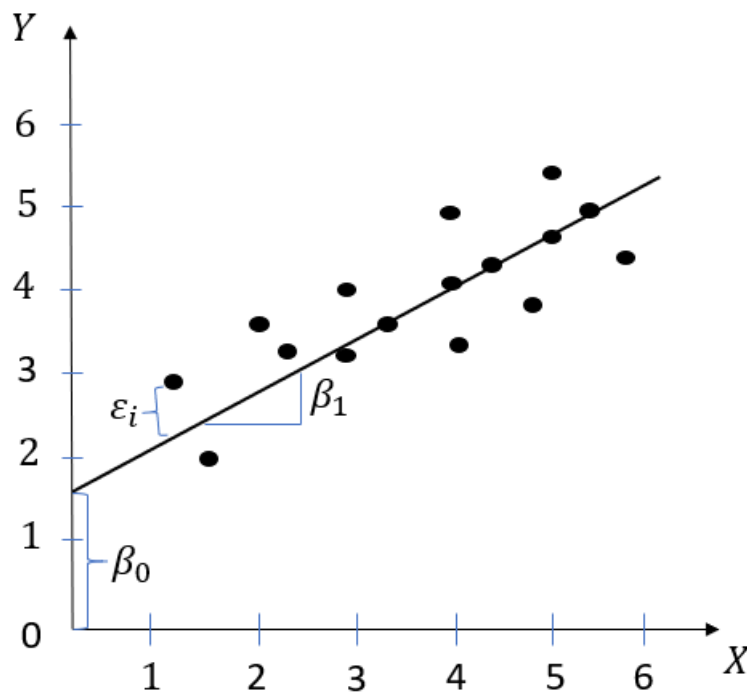


Figura 1. Regresión lineal y su interpretación geométrica

Fuente: Elaborado por el equipo de Trabajo

CONDICIONES PARA LA REGRESIÓN LINEAL MÚLTIPLE

Los modelos de regresión lineal múltiple requieren de las mismas condiciones que los modelos lineales simples más otras adicionales.

No colinealidad o multicolinealidad



En los modelos lineales múltiples los predictores deben ser independientes, no debe haber colinealidad entre ellos. La colinealidad ocurre cuando un predictor está linealmente relacionado con uno o varios de los otros predictores del modelo o cuando es la combinación lineal de otros predictores. Como consecuencia de la colinealidad no se puede identificar de forma precisa el efecto individual que tiene cada una de las variables colineales sobre la variable respuesta, lo que se traduce en un incremento de la varianza de los coeficientes de regresión estimados hasta el punto que resulta prácticamente imposible establecer su significancia estadística. Además, pequeños cambios en los datos provocan grandes cambios en las estimaciones de los coeficientes. Si bien la colinealidad propiamente dicha existe solo si el coeficiente de correlación simple o múltiple entre algunas de las variables independientes es 1, esto raramente ocurre en la realidad. Sin embargo, es frecuente encontrar la llamada *casi- colinealidad o multicolinealidad no perfecta*.

La muestra (y por tanto en la población), ninguna de las variables independientes es constante y no hay una relación lineal constante entre las variables independientes. Una vez que se tiene una muestra de datos, se necesita saber que estos pueden emplearse para calcular las estimaciones de mínimos cuadrados ordinarios, esto quiere decir que en cada variable independiente hay variaciones muestrales y no existe una relación lineal exacta entre las variables independientes, por tanto los parámetros pueden ser calculados (Tobergte & Curtis. 2013).

No existe un método estadístico concreto para determinar la existencia de colinealidad o multicolinealidad entre los predictores de un modelo de regresión, sin embargo, se han desarrollado numerosas reglas prácticas que tratan de determinar en qué medida afecta a la estimación y contraste de un modelo. Los pasos recomendados a seguir son:

- Si el coeficiente de determinación R^2 es alto, pero ninguno de los predictores resulta significativo, hay indicios de colinealidad.
- Calcular una matriz de correlación en la que se estudia la relación lineal entre cada par de predictores. Es importante tener en cuenta que, a pesar de no obtenerse ningún coeficiente de correlación alto, no está asegurado que no exista multicolinealidad. Se puede dar el caso de tener una relación lineal casi perfecta entre tres o más variables y que las correlaciones simples entre pares de estas mismas variables no sean mayores que 0.5.
- Tolerancia (TOL) y Factor de Inflación de la Varianza (VIF). Se trata de dos parámetros que vienen a cuantificar lo mismo (uno es el inverso del otro). El VIF de cada predictor se calcula según la siguiente fórmula:

$$VIF_{\hat{\beta}_j} = \frac{1}{1 - R^2} \quad (2)$$

$$Tolerancia_{\hat{\beta}_j} = \frac{1}{VIF_{\hat{\beta}_j}} \quad (3)$$

donde R^2 se obtiene de la regresión del predictor X_j sobre los otros predictores.

Esta es la opción más recomendada, los límites de referencia que se suelen emplear son:

- ❖ $VIF = 1$: Ausencia total de colinealidad
- ❖ $1 < VIF < 5$: La regresión puede verse afectada por cierta colinealidad.
- ❖ $5 < VIF < 10$: Causa de preocupación
- ❖ El término tolerancia es $1/VIF$ por lo que los límites recomendables están entre 1 y 0.1.

En caso de encontrar colinealidad entre predictores, hay dos posibles soluciones. La primera es excluir uno de los predictores problemáticos intentando conservar el que, a juicio del investigador, está influyendo realmente en la variable respuesta. Esta medida no suele tener mucho impacto en el modelo en cuanto a su capacidad predictiva ya que,

al existir colinealidad, la información que aporta uno de los predictores es redundante en presencia del otro. La segunda opción consiste en combinar las variables colineales en un único predictor, aunque con el riesgo de perder su interpretación.

Cuando se intenta establecer relaciones causa-efecto, la colinealidad puede llevar a conclusiones muy erróneas, haciendo creer que una variable es la causa cuando en realidad es otra la que está influenciando sobre ese predictor

Parsimonia

Este término hace referencia a que el mejor modelo es aquel capaz de explicar con mayor precisión la variabilidad observada en la variable respuesta empleando el menor número de predictores, por lo tanto, con menos asunciones.

Relación lineal entre los predictores numéricos y la variable respuesta

Cada predictor numérico tiene que estar linealmente relacionado con la variable respuesta Y mientras los demás predictores se mantienen constantes, de lo contrario no se puede introducir en el modelo. La forma más recomendable de comprobarlo es representando los residuos del modelo frente a cada uno de los predictores. Si la relación es lineal, los residuos se distribuyen de forma aleatoria entorno a cero. Estos análisis son solo aproximados, ya que no hay forma de saber si realmente la relación es lineal cuando el resto de predictores se mantienen constantes.

El modelo poblacional puede escribirse como:

$$y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$$

Describe la relación poblacional, que se espera estimar y explícitamente establece que existe una relación lineal entre los parámetros k a estimar (Tobergte & Curtis, 2013).

Distribución normal de los residuos

Los residuos se deben distribuir de forma normal con media cero. Para comprobarlo se recurre a histogramas, a los cuantiles normales o a test de hipótesis de normalidad.

Error poblacional u es independiente de las variables explicativas x_1, x_2, \dots, x_k y está distribuido normalmente siendo su media cero y su varianza σ^2 : $u \sim N(0, \sigma^2)$. Se agregó el supuesto de normalidad de manera que puedan realizarse pruebas de hipótesis exactas, puede omitirse si se tiene un tamaño de muestra razonablemente grande. Además, que implica una propiedad de eficiencia más fuerte de los Mínimos Cuadrados Ordinarios (MCO), esto son los que tienen la menor varianza entre todos los estimadores insesgados (Tobergte & Curtis, 2013).

H_0 : Los datos provienen de una distribución normal.

H_a : Los datos no provienen de una distribución normal.

La prueba de Shapiro -Wilk (Shapiro 1965) para determinar la normalidad del set de datos. Dicho estadístico es apropiado tanto para muestras pequeñas (i.e. <50 observaciones) como para muestras grandes (i.e. 1000 o más observaciones). Se le considera como una de las pruebas más poderosas para probar la normalidad.

El estadístico de prueba es.

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (4)$$

Donde:

- $x_{(i)}$ es el número que ocupa la i – ésima posición de la muestra.
- $\bar{x} = (x_1 + \dots + x_n)/n$ es la media muestral
- Las constantes a_1 se calculan con:



$$(\mathbf{a}_1, \dots, \mathbf{a}_2) = \frac{\mathbf{m}^T \mathbf{V}^{-1}}{(\mathbf{m}^T \mathbf{V}^{-1} \mathbf{m})^{\frac{1}{2}}} \quad (5)$$

$$\mathbf{m} = (m_1, \dots, m_n)^T$$

Donde, m_1, \dots, m_n son valores medio del estadístico ordenado, de variables aleatorias independientes e idénticamente distribuidas, V es la matriz de covarianzas de ese estadístico de orden.

Regla de rechazo

Valor aproximado p : Rechaza H_0 si valor $p \leq \alpha$

Valor critico aproximado: Rechaza H_0 si valor $W \geq W_t$

Variabilidad constante de los residuos (homocedasticidad)

La varianza de los residuos debe de ser constante en todo el rango de observaciones. Para comprobarlo se representan los residuos. Si la varianza es constante, se distribuyen de forma aleatoria manteniendo una misma dispersión y sin ningún patrón específico. Una distribución cónica es un claro identificador de falta de homocedasticidad. También se puede recurrir a contrastes de homocedasticidad como el test de Breusch-Pagan.

Las hipótesis a contrastar son:

H_0 : Existe homocedasticidad $\sigma_1^2 = \sigma_t^2 \forall t = 1, \dots, T$

H_a : No existe homocedasticidad

Se calcula la suma de cuadrados de la regresión.

Se calcula el estadístico de este test que viene dado por



$$BP = SCR/2$$

Decisión

Se rechaza H_0 si, el estadístico BP es mayor que el valor tabular.

si $p < \alpha$ se rechaza H_0 , es decir se acepta H_a

No autocorrelación (Independencia)

Los valores de cada observación son independientes de los otros, esto es especialmente importante de comprobar cuando se trabaja con mediciones temporales. Se recomienda representar los residuos ordenados acorde al tiempo de registro de las observaciones, si existe un cierto patrón hay indicios de autocorrelación. También se puede emplear el test de hipótesis de Durbin-Watson.

Arenas (2011). Para fenómenos temporales los datos obtenidos comúnmente presentan ciclos y tendencias, los cuales al realizar una regresión lineal son captados por los residuales del modelo causando que estos estén relacionados con el valor del residual inmediato anterior. A lo anterior se le conoce como residuales correlacionados o auto correlacionados.

Valores atípicos, con alto Leverage o influyentes

Es importante identificar observaciones que sean atípicas o que puedan estar influenciando al modelo. La forma más fácil de detectarlas es a través de los residuos.

Identificación de valores atípicos (outliers), de alto leverage o influyentes

Independientemente de que el modelo se haya podido aceptar, siempre es conveniente identificar si hay algún posible Outlier, observación con alto leverage o influyente, puesto que podría estar condicionando en gran medida el modelo. La eliminación de este tipo de observaciones debe de analizarse con detalle y dependiendo de la finalidad del modelo.



Si el fin es predictivo, un modelo sin Outliers ni observaciones altamente influyentes suele ser capaz de predecir mejor la mayoría de casos. Sin embargo, es muy importante prestar atención a estos valores ya que, de no tratarse de errores de medida, pueden ser los casos más interesantes. El modo adecuado a proceder cuando se sospecha de algún posible valor atípico o influyente es calcular el modelo de regresión incluyendo y excluyendo dicho valor. En el capítulo Regresión Lineal Simple se describe con detalle cómo realizar el análisis para detectarlos.

Yamamoto (2017) indica que la teoría de la detección de Outliers una forma simple para calcular Outliers, es definir el límite como 1,5 veces el rango Inter cuartil, por sobre el tercer cuartil o bajo el primero.

Tamaño de la muestra

No se trata de una condición de por sí, pero, si no se dispone de suficientes observaciones, predictores que no son realmente influyentes podrían parecerlo. En el libro Handbook of biological statistics recomiendan que el número de observaciones sea como mínimo entre 10 y 20 veces el número de predictores del modelo. Es importante esclarecer que la gran mayoría de condiciones se verifican utilizando los residuos, por lo tanto, se suele generar primero el modelo y posteriormente validar las condiciones. De hecho, el ajuste de un modelo debe verse como un proceso iterativo en el que se ajusta el modelo, se evalúan sus residuos y se mejora. Así hasta llegar a un modelo óptimo.

Estimación de los parámetros por mínimos cuadrados ordinarios

Uriel (2013) En el modelo de regresión lineal múltiple, el regresando -que puede ser la variable endógena o una transformación de las variables endógenas-, es una función lineal de k regresores correspondientes a las variables explicativas -o a transformaciones de las mismas- y una perturbación aleatoria o error. El modelo también incluye un término

independiente. Si designamos por y al regresando, por x_2, x_3, \dots, x_k a los regresores y por u al error o perturbación aleatoria, el modelo poblacional de regresión lineal múltiple vendrá dado por la siguiente expresión:

$$y_i = \beta_0 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$$

Donde: $\beta_1 + \beta_2 + \beta_3 + \dots + \beta_k$, son fijos y desconocidos. Supongamos que tenemos una muestra aleatoria de tamaño n , $\{(y_i, X_{2i}, X_{3i}, \dots, X_{ki}) : i = 1, 2, \dots, n\}$ extraída de la población estudiada. Si expresamos el modelo poblacional para todas las observaciones de la muestra, se obtiene el siguiente el sistema:

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{k1} + \mu_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{k2} + \mu_2$$

$$y_3 = \beta_0 + \beta_1 x_{31} + \beta_2 x_{32} + \dots + \beta_k x_{k3} + \mu_3$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{1n} + \beta_2 x_{2n} + \dots + \beta_k x_{kn} + \mu_n$$

Se trata entonces de un modelo que es representable de manera matricial así:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1}, X = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & x_{32} & \dots & x_{k2} \\ 1 & x_{13} & x_{23} & x_{33} & \dots & x_{k3} \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} & \dots & x_{kn} \end{bmatrix}_{n \times k}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix}_{n \times 1} \text{ y } \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_n \end{bmatrix}_{n \times 1}$$

Donde la variable dependiente es una variable aleatoria que está en función de las k variables independientes y un término de error μ . El primer subíndice de las variables explicativas representa la posición de la variable como tal, mientras que el segundo subíndice representa las n observaciones determinadas de la muestra. Esta forma matricial de escribir el modelo es equivalente a:

$$Y_i = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + \mu_i, i = 1, 2, \dots, n$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}_{n \times 1} = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & x_{32} & \dots & x_{k2} \\ 1 & x_{13} & x_{23} & x_{33} & \dots & x_{k3} \\ & & & \vdots & & \\ & & & \vdots & & \\ & & & \vdots & & \\ 1 & x_{1n} & x_{2n} & x_{3n} & \dots & x_{kn} \end{bmatrix}_{n \times k} * \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{bmatrix}_{n \times 1} + \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \vdots \\ \mu_n \end{bmatrix}_{n \times 1}$$

Se puede también abreviar de la siguiente manera:

$$Y = X\beta + \mu \quad (6)$$

Cálculo de los coeficientes

Al igual que en el modelo de regresión lineal simple, se utiliza el método de Mínimos Cuadrados Ordinarios (MCO), el cual estima la recta obteniendo el Mejor Estimados Lineal Insesgado (MELI) el cual minimiza la suma de los cuadrados de los residuos. El criterio considera que la función que mejor se ajusta a los datos es la que minimiza la varianza del error u , lo que es equivalente a minimizar:

$$S(\beta_0, \beta_1, \beta_2, \dots, \beta_k) = \sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k))^2$$

Para el cálculo de los coeficientes β 's de manera manual, se va a trabajar con la matriz de las X 's, y es importante tener en cuenta, desde el principio, que la matriz $X^T X$ debe ser invertible. Para esto, es necesario que el rango de la matriz X , que es el número de regresores k , sea menor que el número de observaciones n . Este es un primer paso que debe verificar el econométrista antes de empezar a trabajar, aunque por lo general, como se trabaja con muestras representativas, estas son generalmente grandes, y por tanto se cumple con esta condición.

En el modelo de k variables explicativas, los estimadores de MCO se obtienen al reducir:

$$\sum \mu_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \dots - \hat{\beta}_k X_{ki})^2$$

En notación matricial, lo anterior equivale entonces a realizar la reducción de $\hat{\mu}^T \hat{\mu}$ pues:

$$\hat{\mu}^T \hat{\mu} = [\hat{\mu}_1 \quad \hat{\mu}_2 \quad \dots \quad \hat{\mu}_n] \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \vdots \\ \hat{\mu}_n \end{bmatrix} = \hat{\mu}_1^2 + \hat{\mu}_2^2 + \dots + \hat{\mu}_n^2 = \sum \hat{\mu}_i^2$$

Ahora, como el modelo estimado escrito en términos matriciales es: $Y = X\hat{\beta} + \mu$, lo que tenemos es que: $\mu = Y - X\hat{\beta}$

Por lo tanto, de lo anterior se puede deducir fácilmente que:

$$\begin{aligned} \hat{\mu}^T \hat{\mu} &= (Y - X\hat{\beta})^T (Y - X\hat{\beta}) \\ &= Y^T Y - 2\hat{\beta}^T X^T Y + \hat{\beta} X^T X \hat{\beta} \end{aligned}$$

Si el lector conoce las propiedades de la transpuesta de una matriz, debe saber que:

$(X\hat{\beta})^T = \hat{\beta}^T X^T$ y, además, como $\hat{\beta}^T X^T$ es un escalar, entonces es igual a su transpuesta $Y^T X \hat{\beta}$.

El método de MCO consiste pues, como ya se mencionó en el documento previo a este, en estimar unos coeficientes β de tal manera que se minimice $\sum \hat{\mu}_i^2$

Como lo expone Gujarati (2010) de este proceso resultan k ecuaciones simultáneas con k incógnitas, que son las ecuaciones normales de la teoría de mínimos cuadrados. Dichas ecuaciones escritas de forma matricial son:

$$\begin{matrix}
 \begin{bmatrix} n & \sum X_{2i} & \dots & \sum X_{ki} \\ \sum X_{2i} & \sum X_{2i}^2 & \dots & \sum X_{2i} X_{ki} \\ \vdots & \vdots & \ddots & \vdots \\ \sum X_k & \sum X_{2i} X_{ki} & \dots & \sum X_{ki}^2 \end{bmatrix} & \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_{ki} \end{bmatrix} & = & \begin{bmatrix} 1 & 1 & 1 & 1 \\ X_{2i} & X_{22} & \dots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{ki} & X_{x2} & \dots & X_{kn} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \\
 X^T X & \hat{\beta} & & X^T & Y
 \end{matrix}$$

Es decir;

$$(X^T X)\hat{\beta} = X^T Y$$

Si se multiplica a ambos lados de la ecuación por $(X^T X)^{-1}$ con el fin de hallar la matriz de los β 's se obtiene entonces que:

$$(X^T X)^{-1}(X^T X)\hat{\beta} = (X^T X)^{-1}(X^T Y)$$

Donde $(X^T X)^{-1}(X^T X)$ es la matriz identidad, y por tanto se ha encontrado el cálculo para hallar el vector de coeficientes β así:

$$\hat{\beta} = (X^T X)^{-1}(X^T Y)$$

Se sabe además que β es una matriz de $(k \times 1)$, $(X^T X)^{-1}$ es una matriz de $(k \times k)$, X^T es una matriz de $(k \times n)$ y Y de $(n \times 1)$.

El cálculo de los estimadores de MCO para el modelo de regresión lineal múltiple es entonces posible realizarlo manualmente hallando la matriz inversa de $(X^T X)$ mediante métodos como el de Gauss-Jordan o el método de cofactores, por ejemplo, y posteriormente realizando los cálculos de multiplicación de matrices básicos de un curso de Álgebra matricial.

El método de MCO arroja los estimadores β insesgados y de mínima varianza:

1. Insesgamiento: el valor esperado de la estimación es aproximadamente igual al valor del parámetro poblacional, es decir: $E(\hat{\theta}) \approx \theta$.
2. Mínima Varianza: los coeficientes estimados son consistentes, lo cual se muestra mediante una varianza tal que: $s_y^2 = var(y) = \frac{\sigma^2}{n}$

(Espitia). El término independiente β_0 no necesariamente debe aparecer en el modelo, sin embargo, es la muestra de una variable X_0 cuyo valor siempre es 1.

El objetivo de la estimación es obtener los valores de los parámetros $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ del modelo los cuales, una vez estimados, se representan por $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k$.

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \dots + \hat{\beta}_k x_{ki} \quad i = 1, 2, \dots, n$$

Hipótesis parcial individual

Suponiendo que los errores tienen una distribución $N(0, \sigma^2)$ podemos utilizar la prueba t para demostrar que un coeficiente parcial de regresión individual, cuya hipótesis nula sería la variable X_i estando todas las demás variables constantes tiene efectos en la variable Y , con $n-k$ grados de libertad a dos colas y con un valor alfa de significancia.

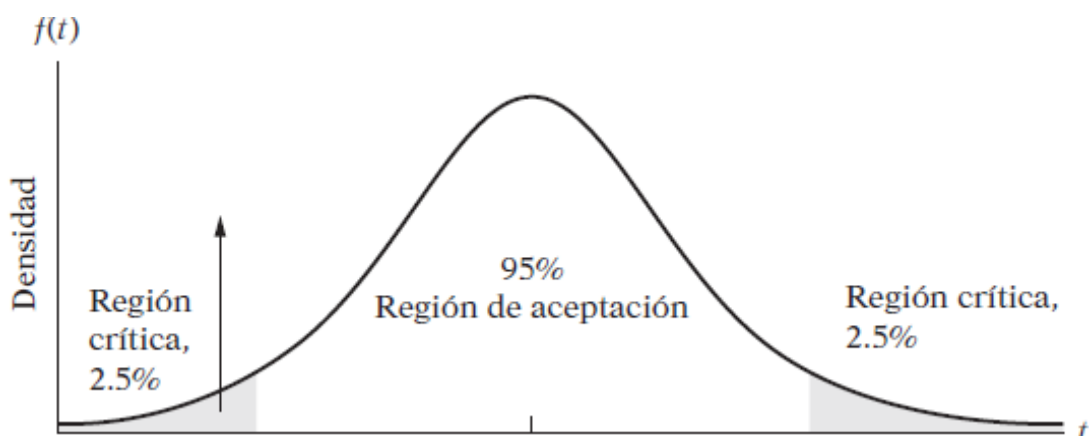


Figura 2. Prueba de contraste de la hipótesis "t"

Fuente: Tobergte & Curtis (2013)



Test Fisher: F prueba de significancia de la regresión

En el modelo de regresión lineal múltiple prueba la significancia conjunta de todas las variables independientes

La prueba de hipótesis del test F es:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_a : \beta_1 \neq \beta_0 \neq \dots \neq \beta_k \neq 0$$

El F no prueba la significancia de β_0

$$F = \frac{\frac{SCE}{gl}}{\frac{SCR}{n-k}} = \frac{SCE}{k-1} \cdot \frac{n-k}{SCR} \quad (7)$$

Donde:

SCE : Suma de Cuadrados de los errores.

SCR : Suma de cuadrados de la regresión.

k: Número de variables independientes.

n: Número de observaciones

gl: Grados de libertad.

- $F_{\alpha}(k-1, n-k)$ es el valor F crítico en el nivel de significancia α
- $(k-1)gl$ en el numerador y $(n-k)gl$ en el denominador.

Regla de rechazo.

Valor aproximado p : Rechaza H_0 si valor $p \leq \alpha$

Valor critico aproximado: Rechaza H_0 si valor $F \geq F$

Mean Square Error - Error Medio Cuadrático (MSE)

Cada error o residual se eleva al cuadrado; luego, estos valores se suman y se divide entre el mismo número de observaciones. Este enfoque penaliza los errores mayores de pronóstico ya que eleva cada uno al cuadrado. Esto es importante pues en ocasiones pudiera ser preferible una técnica que produzca errores moderados a otra que por lo regular tenga errores pequeños, pero que ocasionalmente arroje algunos en extremo grandes. Idealmente, se desea que el resultado del EMC sea 0, lo que indicaría un pronóstico perfecto. En general, mientras más pequeño sea el valor de EMC, mejor es el modelo de pronóstico (Ramírez, 2013).

$$MSE = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}$$

Raíz cuadrada del error medio cuadrático (RMSE)

La sensibilidad del RMSE a valores atípicos es la característica más común con el uso de esta métrica, Chai & Draxler (2014).

De hecho, la existencia de valores atípicos y su probabilidad de ocurrencia está bien descrita por la distribución normal subyacente al uso de la métrica RMSE cuya fórmula es la siguiente:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n}}$$

El error estándar

La variabilidad de las medias muestrales se puede medir por su desviación estándar. Esta medida se conoce como el *error estándar* y tiende a disminuir cuando aumenta el tamaño de la(s) muestra(s).

Cuanto menor sea el error estándar, más precisa será la estimación.



Correlación

La correlación es un tipo de asociación entre dos variables numéricas, específicamente evalúa la tendencia (creciente o decreciente) en los datos.

- ❖ un valor positivo indica una relación directa o positiva.
- ❖ un valor negativo indica relación indirecta, inversa o negativa.
- ❖ un valor nulo indica que no existe una tendencia entre ambas variables (puede ocurrir que no exista relación o que la relación sea más compleja que una tendencia, por ejemplo, una relación en forma de U).

Toma valores entre -1 a 1. Cuanto más cercano sea el valor a los extremos del intervalo (1 o -1) más fuerte será la tendencia de las variables, o será menor la dispersión que existe en los puntos alrededor de dicha tendencia. Cuanto más cerca del cero esté el coeficiente de correlación, más débil será la tendencia, es decir, habrá más dispersión en la nube de puntos.

Si la correlación vale 1 o -1 diremos que la correlación es “perfecta”,

Si la correlación vale 0 diremos que las variables no están correlacionadas.

Bondad de ajuste

Vallejo (2012) afirma que cuán “bien” se ajusta la línea de regresión a los datos es claro que, si todas las observaciones cayesen en la línea de regresión, obtendríamos un ajuste “perfecto”, pero rara vez se presenta este caso. Por lo general hay algunas u_i positivas y algunas u_i negativas. Se tiene la esperanza de que estos residuos alrededor de la línea de regresión sean lo más pequeños posibles cuyos resultados son medidos generalmente por un valor llamado R^2 .

Gordón & Camargo (2015) determina que el R^2 mide la proporción (porcentaje si lo multiplicamos por 100) de la variabilidad de la variable dependiente explicada por las

variables independiente que en ese momento han sido admitidas en el modelo. A partir del resumen de los modelos generados paso a paso podemos calcular el incremento de R^2 (Ramírez, 2013) mide el porcentaje del cambio de una variable dependiente explicado por el cambio de las variables independientes, a través de un modelo de regresión lineal múltiple. Entre más cercano a uno mejor es el ajuste. Descrito en mejor detalle en las referencias

La fórmula queda definida de la siguiente manera:

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$$

Donde:

\hat{Y}_i : Es el valor estimado en la regresión para la variable Y.

Y_i : Es el valor real de la variable Y o valor observado.

\bar{Y} : Es el promedio de la variable dependiente Y.

Criterios de aceptación.

1. El grado de ajuste de una función es aceptable cuando $R^2 \geq 0.8$
2. Sí varias funciones cumplen con esta condición se elige la que se acerca más a 1.

2.2.2. Árboles de regresión

Los métodos predictivos como la regresión lineal o polinómica generan modelos globales en los que una única ecuación se aplica a todo el espacio muestral. Cuando el estudio implica múltiples predictores, que interaccionan entre ellos de forma compleja y no lineal, es muy difícil encontrar un modelo global que sea capaz de reflejar la relación entre las variables. Existen métodos de ajuste no lineal (step functions, splines,...) que



combinan múltiples funciones y que realizan ajustes locales, sin embargo, suelen ser difíciles de interpretar. Los métodos estadísticos basados en árboles engloban a un conjunto de técnicas supervisadas no paramétricas que consiguen segmentar el espacio de los predictores en regiones simples, dentro de las cuales es más sencillo manejar las interacciones.

Los métodos basados en árboles se han convertido en uno de los referentes dentro del ámbito predictivo debido a los buenos resultados que generan en ámbitos muy diversos.

Árboles de predicción en R

Existen diferentes paquetes en R que permiten la creación de modelos basados en árboles:

- tree y rpart: contiene funciones para la creación y representación de árboles de regresión y clasificación.
- rpart.plot: permite crear representaciones detalladas de modelos creados con rpart.
- randomForest: implementa algoritmos para crear modelos random forest.
- gbm: implementa algoritmos de boosting.
- C50: contiene los algoritmos C5.0 para árboles de clasificación.

Predicción

Tras la creación de un árbol, las observaciones de entrenamiento quedan agrupadas en los nodos terminales. Para predecir una nueva observación, se recorre el árbol en función de los valores que tienen sus predictores hasta llegar a uno de los nodos terminales. En el caso de regresión, el valor predicho suele ser la media de la variable respuesta de las observaciones de entrenamiento que están en ese mismo nodo. Si bien la media es el valor más empleado, se puede utilizar cualquier otro (mediana, cuantil...).

Existen principalmente tres procedimientos: CHAID (Chi-Square Automatic Interaction Detector), QUEST (Quick Unbiased Efficient Statistical Tree) y CART, en el que nos centraremos por ser relativamente sencillo y no fácil de implementar en R.

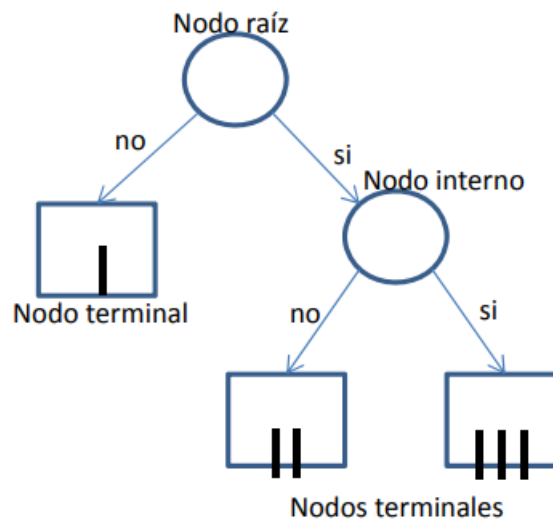


Figura 3. Elementos del Árbol de regresión

Fuente: Elaboración a partir la revisión literaria.

Como se muestra en la figura, el árbol tiene tres niveles de nodos, el primer nivel tiene un único nodo en la cima llamado *nodo raíz*. Un *nodo interno* en el segundo nivel, y tres *nodos terminales* que están respectivamente en el segundo y tercer nivel. El nodo raíz y el nodo interno son particionados cada uno en dos nodos en el siguiente nivel los cuales son llamados nodos hijos (o ramas) izquierdo y derecho.

Heredia, B & B (2010). Define los árboles de clasificación y regresión, conocidos como algoritmos CART (del inglés classification and regression trees), constituyen una aproximación multivariada no paramétrica que permite identificar y dimensionar las variables X de mayor impacto en una variable Y. Los modelos CART particionan los datos en forma recursiva de modo tal de conformar subconjuntos cada vez más homogéneos en base a criterios de partición de las variables explicativas. Cada árbol se obtiene a partir de la clasificación de un nodo parental o raíz que contiene la totalidad

de los datos, mediante un algoritmo de partición especificado en función de un criterio de partición referido al tamaño del nodo formado o a la variabilidad contenida en los datos del nodo, también el algoritmo CART. Un árbol de regresión o de clasificación consiste en un conjunto de reglas determinadas por un procedimiento de ajuste mediante particiones binarias recursivas.

Todos los objetos son considerados como pertenecientes al mismo grupo. En cada instancia de separación el algoritmo analiza todas las variables regresoras y selecciona, para realizar la partición binaria de los datos, aquella que permite conformar dos subgrupos o nodos más homogéneos dentro y más heterogéneos entre ellos.

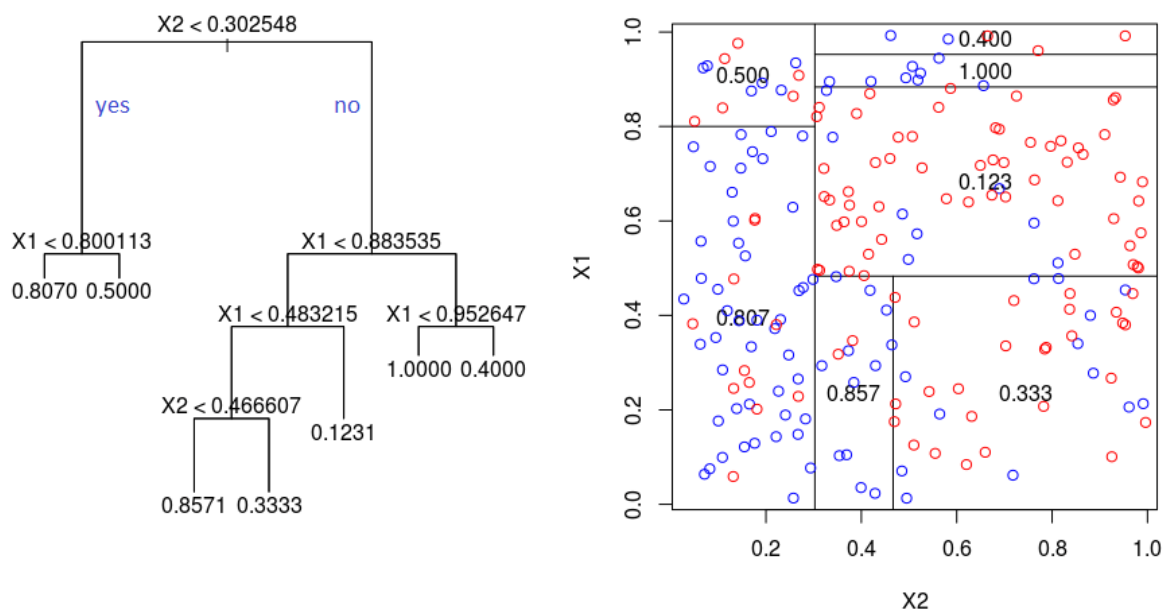


Figura 4. Particiones del algoritmo CART

Fuente: Heredia, Bruno & Balzarini (2010).

De la figura 03; los árboles de regresión permiten predecir variables respuestas continuas, las observaciones según umbrales de las variables regresoras, considerando la

suma de cuadrados de la respuesta como medida de heterogeneidad dentro de cada partición. Como en el árbol de clasificación, la medida de heterogeneidad entre las observaciones que quedan dentro de un nodo debe ser menor que la calculada entre las observaciones de distintos nodos.

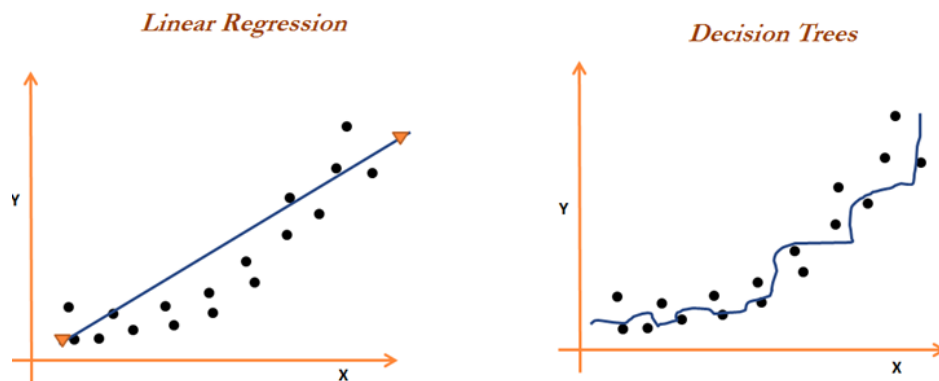


Figura 5. Ilustración gráfica de la Regresión lineal y el Árbol de decisión

Fuente: Heredia, Bruno & Balzarini (2010).

Podado del árbol

El árbol obtenido es generalmente sobreajustado por tanto es podado, cortando sucesivamente ramas o nodos terminales hasta encontrar el tamaño “adecuado” del árbol (Pineda, 2009).

Con la finalidad de reducir la varianza del modelo y así disminuir el *test error*, se somete al árbol a un proceso de *poda*, intenta encontrar el árbol más sencillo (menor tamaño) que consigue explicar las observaciones, y así para identificar el valor óptimo de penalización α . Por defecto, esta función emplea la desviación para guiar el proceso de *poda* y manteniendo la estructura robusta que consigue un test error bajo. La selección del sub-árbol óptimo puede hacerse mediante la validación cruzada, sin embargo, dado



que los árboles se crecen lo máximo posible (tienen muchos nodos terminales) no suele ser viable estimar el test error de todas las posibles sub-estructura que se pueden generar.

2.3. MARCO CONCEPTUAL

2.3.1. Volumen

El volumen es la cantidad de acciones o contratos negociados en un período de tiempo determinado. En los gráficos se lo puede clasificar de acuerdo al tiempo los volúmenes diarios, semanales, mensuales, etc. A mayor volumen operado, mayor interés sobre ese activo. Y a menor volumen operado, menor interés sobre el activo (Eainversores, 2020).

2.3.2. Precio Máximo

Es el precio máximo que puede alcanzar una acción de una corporación u empresa en un periodo de tiempo determinado, monetizado en la unidad monetaria denominada como dólar

2.3.3. Precio de Apertura

En el presente trabajo de investigación, el precio de apertura es aquel con el que se apertura al mercado a la venta en un determinado periodo de tiempo el cual esta monetizado en dólares estadounidenses.

2.3.4. Acciones Ordinarias

Ayala & Samillán (2019) define las acciones como partes iguales en las que se divide el capital social de una sociedad anónima. Estas partes son poseídas por una persona, que recibe el nombre de accionista, y representan la propiedad que la persona tiene de la empresa, es decir, el porcentaje de la empresa que le pertenece al accionista.



CAPÍTULO III

MATERIALES Y MÉTODOS

3.1. TIPO Y DISEÑO DE INVESTIGACIÓN

3.1.1. Tipo de investigación

El tipo de investigación es cuantitativo descriptivo correlacional, ya que se describe las variables estudiadas y correlacional porque busca la relación entre los modelos de Regresión lineal múltiple y Árbol de regresión

3.1.2. Diseño de investigación

Una investigación es de tipo no experimental puesto que los datos fueron recolectados previamente (Fernández & Baptista, 2014).

El diseño de investigación para este estudio es no experimental y de corte transversal por tanto no se manipula las variables además los datos fueron obtenidos en un corte de tiempo.

3.2. UBICACIÓN DEL TRABAJO DE INVESTIGACIÓN

El presente trabajo de investigación se realizó en a la ciudad de Puno, Provincia de Puno departamento de Puno a los 3852 msnm.

3.3. POBLACIÓN Y MUESTRA

3.3.1. Población

La población para el estudio está estuvo conformada todas las Acciones de la Empresa Intel a través del tiempo, desde su creación en la bolsa de valores por tanto es una población indeterminada o infinita.

3.3.2. Muestra

La muestra para el trabajo de investigación es de tipo no probabilística obtenida por conveniencia, por ende, se consideró a todas las acciones de la empresa Intel las mismas



que se vendieron desde mayo del 2018 hasta octubre del 2019 lo que genera una suma total de 410 registros recopilados.

3.4. ALCANCE DE LA INVESTIGACIÓN

El alcance de la investigación es de tipo predictivo. La misma que tiene como propósito prever o anticipar situaciones futuras, requiere de la exploración, además de la descripción, la comparación, el análisis y la explicación. La investigación de tipo pronóstico es aquella en la cual el propósito principal es “predecir” la dirección futura de los eventos investigados. Sampieri et al. (1998).

El alcance de esta investigación se realiza a la empresa Intel, de este modo servirá para los accionistas a reducir la incertidumbre al momento de vender una acción en un determinado periodo de tiempo, utilizando la Regresión lineal múltiple y los Árboles de regresión, se optará aquella que presente menor error, y R^2 alta, la predicción se dará por día, mas no a largo plazo.

Variable dependiente (Y)

El precio máximo de venta que

la acción ha logrado durante la negociación regular por un determinado periodo de tiempo; durante el día, para esta variable la unidad de medida es el dólar (\$).

Variables independientes (X's)

- ❖ X_1 : Precio de apertura en el cual la acción es abierta al mercado (esta varia a lo largo del día). Al igual que la variable anterior, la unidad de medida es el dólar (\$).
- ❖ X_2 : Volumen de ventas de un determinado periodo de tiempo de la empresa Intel. Para esta variable la unidad de es la cantidad de ventas que se registró, durante el día.



Unidad muestral

En el estudio la unidad muestral fueron las Acciones de la Empresa Intel 2019.

Unidad de medida

En la investigación, la unidad de medida es el dólar debido a que el precio de una acción en el mercado estadounidense se da en dólares (\$).

3.5. TÉCNICA E INSTRUMENTOS DE RECOLECCIÓN DE DATO

3.5.1. Técnica

Torres & Paz (2019), indican que, las fuentes de Información, son todos aquellos medios de los cuales procede la información, que satisfacen las necesidades de conocimiento de una situación o problema presentado, que posteriormente será utilizado para lograr los objetivos esperados, las fuentes secundarias son aquellas que parten de datos pre-elaborados, como pueden ser datos obtenidos de anuarios estadísticos, de Internet, de medios de comunicación.

Para el estudio se hizo la obtención de datos mediante el internet, los cuales han sido obtenidos de la página <https://www.nasdaq.com/es/market-activity/stocks/intc> , “Nasdaq” puesto que la técnica utilizada fue documental. Para ello se recurre a la fuente secundaria electrónica.

3.5.2. Instrumento:

Para llevar a cabo la investigación, se acudió a la base de datos electrónica; como instrumento de recolección, ver anexo 2.

3.6. METODOLOGÍA ESTADÍSTICA UTILIZADA (procedimiento)

Se utilizó la metodología propuesta por Kleinbaum & Kupper (1998). Fue útil para realizar la comparación de las métricas estadísticas utilizadas para esta investigación.

Obtención de variables



En esta etapa se realiza lo siguiente

- Obtención de la base de datos electrónica “Nasdaq”, a partir siguiente Link <https://www.nasdaq.com/es/market-activity/stocks/intc>, se efectuó la obtención de la muestra.
- Selección de variables de interés.
- Se toma las variables independientes y la variable dependiente.
- Se realiza la limpieza de datos.

3.7. ELABORACIÓN DE REGRESIÓN LINEAL MÚLTIPLE

Una vez definida las variables, Y y X' s, dependiente e independientes, se realiza la preparación de la data; para exportar al paquete estadístico R, se realiza la obtención del Primer, Segundo y Tercer modelo de Regresión lineal múltiple, a su vez se analiza los resultados correspondientes para cada modelo de regresión.

Elección de parámetros para la regresión lineal múltiple.

En esta etapa se valida los resultados del modelo de Regresión lineal múltiple y la verificación de los supuestos, así como el R^2 , el Estadístico F y el promedio del error estándar.

3.8. ELABORACIÓN DEL ARBOL DE REGRESIÓN

Para la elaboración del Árbol de regresión se dispone de los datos,

Entrenamiento:

$$\frac{80}{100} * 410 = 328, \text{ Se utilizó el 80\% del total de datos}$$

Validación:

$$\frac{20}{100} * 410 = 82, \text{ Para luego predecir con el 20\%}$$

los mismos que se exportaron al Paquete R, para tal efecto se realizó la selección aleatoria, siendo 328 observaciones para entrenamiento y 82 para la validación, de esta



forma asegurando que la distribución de las variables en entrenamiento y validación sea la misma.

Calibración del Árbol de Regresión

- Se ajustó el Árbol de regresión y profundidad de nodos para evitar el sobreajuste.
- La obtención del Árbol de regresión.
- Se realizó la poda del Árbol de regresión para obtener mejores resultados.
- Se obtuvo la comparativa del Árbol de regresión sin podado y podado se optó por el menor MSE.

3.9. COMPARACIÓN DE ERRORES

Finalmente se realizó la comparativa entre Regresión lineal múltiple y Árbol de regresión, utilizando las métricas correspondientes y de esta forma se determinó entre Regresión lineal múltiple y el Árbol de regresión.

El intervalo de confianza para esta investigación fue:

- ❖ Nivel confianza $1 - \alpha = 0.95 \approx 95\%$
- ❖ Nivel de significancia $\alpha = 0.05 \approx 5\%$
- ❖ Para Probabilidad $\rho = 0.05$

CAPÍTULO IV

RESULTADOS Y DISCUSIÓN

3.1. ANÁLISIS DESCRIPTIVO

Luego de obtener los datos se procede a realizar análisis descriptivo con la finalidad de observar el comportamiento descriptivo de las variables en estudio.

Tabla 1. Estadísticos descriptivos de las variables en estudio.

Ítem	Volumen	Precio apertura	Precio máximo
Mínimo	6313179	\$ 43.35	\$ 44.36
Primer cuartil	17105165	\$ 47.12	\$ 47.74
Mediana	21336425	\$ 49.5	\$ 50.09
Media	23670816	\$ 50.89	\$ 51.42
Tercer cuartil	27863138	\$ 53.84	\$ 53.84
Máximo valor	86449630	\$ 67.63	\$ 69.29

Fuente. Elaboración por el equipo de trabajo, resultados obtenidos a través del software R.

La tabla 1 muestra los estadísticos descriptivos de las variables en estudio, tal es así que el máximo valor del precio de apertura es de \$ 67.63 y el mínimo \$ 43.35 con un valor promedio \$ 50.89, el valor más alto del precio máximo es de \$ 69.29 y el mínimo valor es de \$ 44.36, posteriormente se muestra su promedio de \$51.42. En cuanto al volumen, el número mínimo de transacciones de ventas fue de 6313179 y el máximo valor fue de 86449630 transacciones, y finalmente se muestra el promedio de transacciones de 23670816.

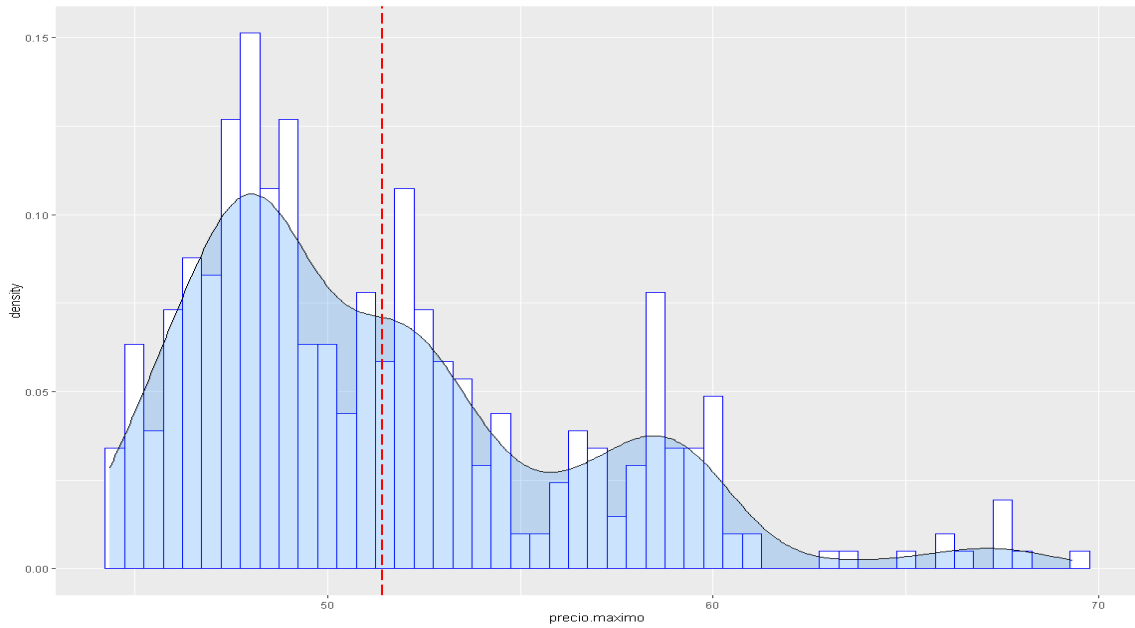


Figura 6. Distribución de la variable precio máximo

Fuente: Gráfico obtenidos a través del software R.

Se observa que la distribución de la variable precio máximo tiene un sesgo a los valores mínimos. rondando la media entre \$ 51.42 dólares, lo cual surge una posible falta de normalidad en esta variable.

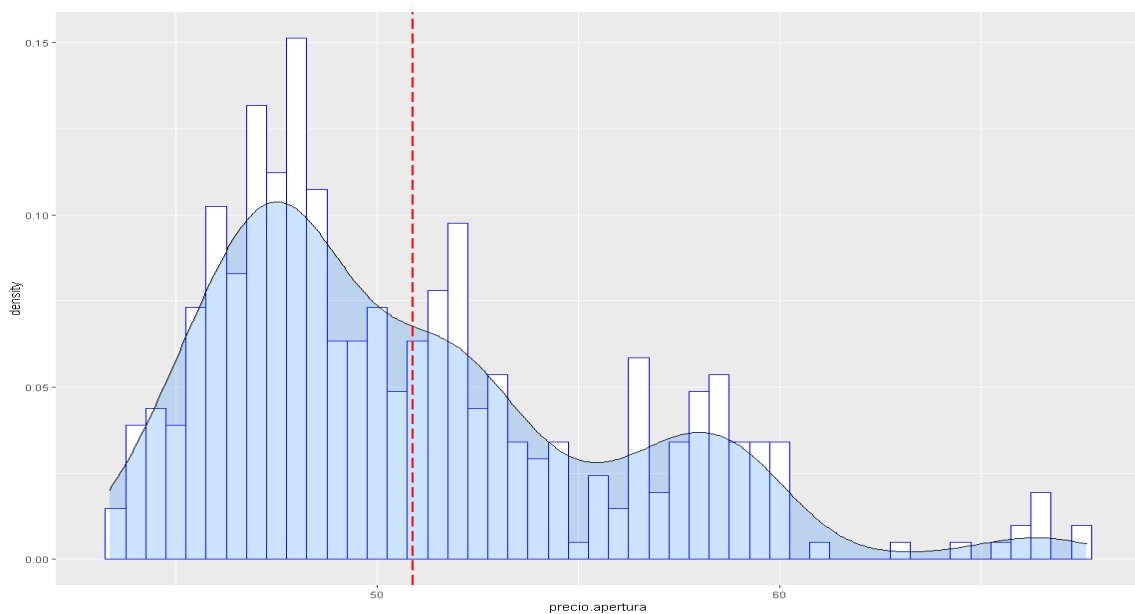


Figura 7. Distribución en la variable precio de apertura.

Fuente: Gráfico obtenidos a través del software R.

De la Figura 7, se observa la distribución de la variable precio de apertura el cual esta sesgado a los valores más pequeños y su media es \$ 50.89 dólares, pero con algunos valores atípicos por encima de los \$ 60 dólares.

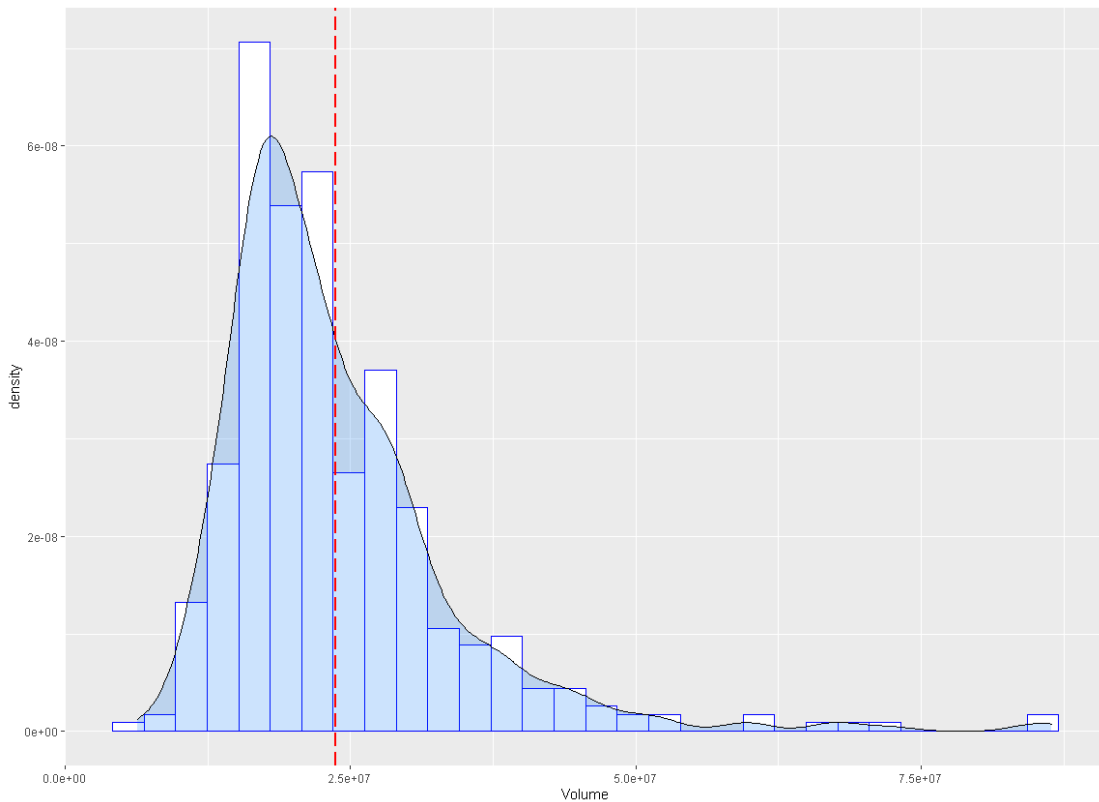


Figura 8. Distribución en la variable volumen

Fuente: Gráfico obtenidos a través del software R.

En la figura 8, se observa un sesgo en la variable volumen, pero con una aparente normalidad, además con una media de 23,670,816 transacciones.

A continuación, se muestra la tabla de correlaciones de las variables Precio de apertura, precio máximo, y volumen.

Tabla 2. *Correlaciones de las variables Precio máximo, Precio de apertura y Volumen a partir de Intel 2019.*

Ítem	Volumen	Precio apertura	Precio máximo
Volumen	1	-0.193	-0.162
precio apertura		1	0.996
precio máximo			1

Fuente. Elaboración por el equipo de trabajo, resultados obtenidos a través del software R.

En la tabla 2, se muestra la presencia de una correlación alta entre la variable precio de apertura y precio máximo con un coeficiente de correlación de 0.996, indicando que, a mayor precio de apertura, mayor precio máximo. La correlación entre la variable precio máximo y volumen es negativa con un valor de -0.162 al igual que la correlación entre la variable precio de apertura y la variable volumen que muestran un coeficiente de correlación de -0.193, esto indica que existe una correlación inversa entre estas variables y el volumen.

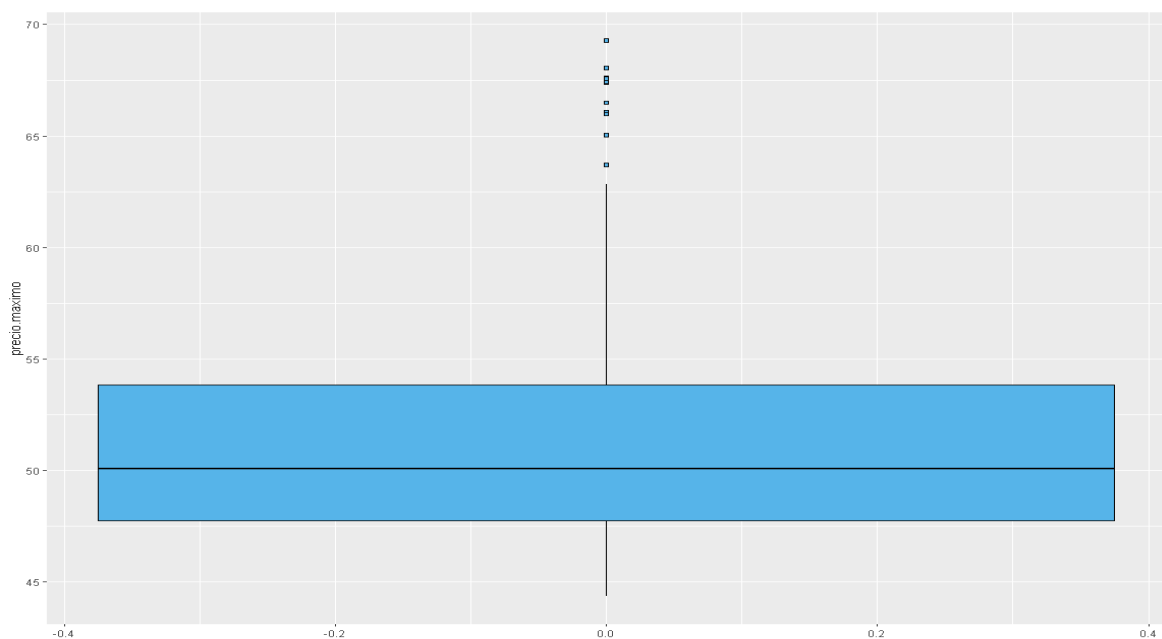


Figura 9. Gráfico de cajas de la variable precio máximo de las acciones de Intel 2019

Fuente: Gráfico obtenidos a través del software R. (ver anexo 2)

En el diagrama de cajas de la variable precio máximo se observa valores atípicos que estarían disminuyendo los valores medios y posiblemente afecten al modelo.

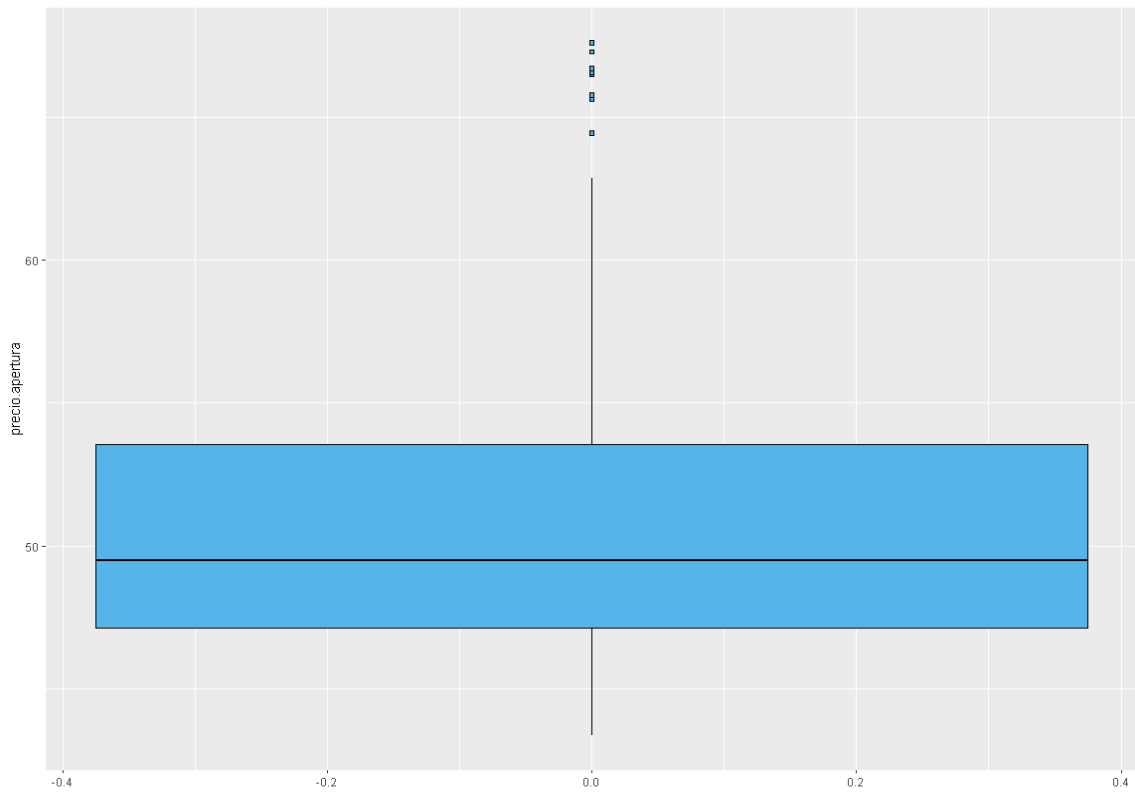


Figura 10. Gráfico de cajas de la variable Precio Máximo de las Acciones de Intel 2019

Fuente: Gráfico obtenidos a través del software R (ver anexo 2).

Igualmente, en el diagrama de cajas de la variable precio de apertura se observa valores atípicos que afectarían al modelo

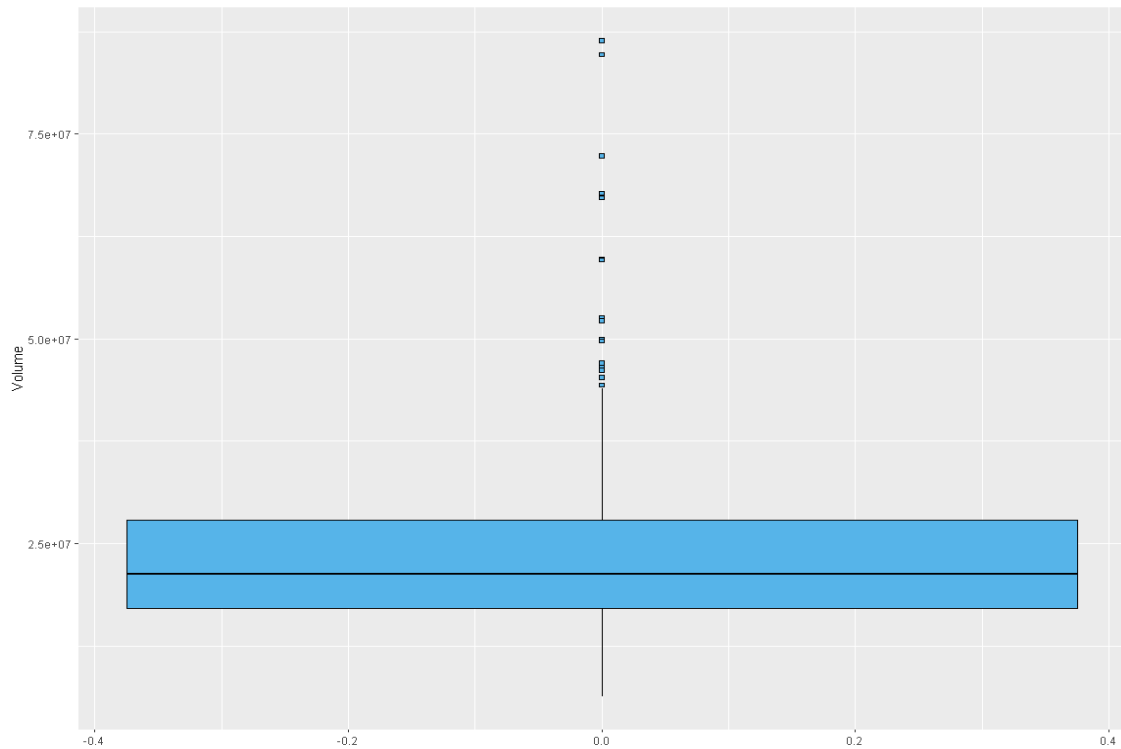


Figura 11. Gráfico de cajas de la variable Volumen de ventas de las acciones de Intel.

Fuente: Gráfico obtenidos a través del software R (ver anexo 02).

La figura 11 muestra el diagrama de cajas para la variable volumen, aparenta ser simétrico puesto que tiene poca variabilidad con mayor concentración de la varianza por encima de la mediana 21,336,425 de transacciones por día.

3.2. MODELO DE REGRESIÓN LINEAL MÚLTIPLE

3.2.1. Obtención del primer modelo de RLM

Para obtener el modelo de regresión lineal múltiple se hace uso del método de los mínimos cuadrados ordinarios.

A continuación, se muestra la salida del paquete R Estadístico.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
Intercepto	1.454e-01	2.238e-01	0.650	0.516
precio.apertura	1.001e+00	4.104e-03	243.841	< 2e-16 ***
volumen	1.492e-08	2.002e-09	7.451	5.6e-13 ***

Residual standard error: 0.4072 on 407 degrees of freedom

Multiple R-squared: 0.9934, Adjusted R-squared: 0.9933

F-statistic: 3.054e+04 on 2 and 407 DF, p-value: < 2.2e-16

Fuente: Resultados obtenidos a través del software R (ver anexo 2).

A partir de los 410 de registros recopilados a partir del Nasdaq, R nos muestra los coeficientes de β_0 , β_1 y β_2 para el primer modelo y las estimaciones correspondientes.

El modelo de regresión lineal múltiple, queda definido como:

$$Y = 0.1454 + 1.00100X_1 + 0.00000001492X_2 \quad (8)$$

Las dos variables independientes precio de apertura y volumen resultan ser significativos $p < \alpha$ (0.05) excepto el intercepto. El modelo en su conjunto observando la prueba F es significativa y la bondad de ajuste es bastante alta $R^2 = 0.9934$. lo que indica que, es capaz de explicar el 99,34% de la variabilidad Observada en el precio máximo. El *p-valué* del modelo es significativo 2.2e-16 por lo que, se puede aceptar que el modelo no es por azar, al menos uno de los coeficientes parciales es distinto de 0.

Referencia

(***). Indica la alta Significancia.

Comprobación de supuestos.

Linealidad de las variables independientes

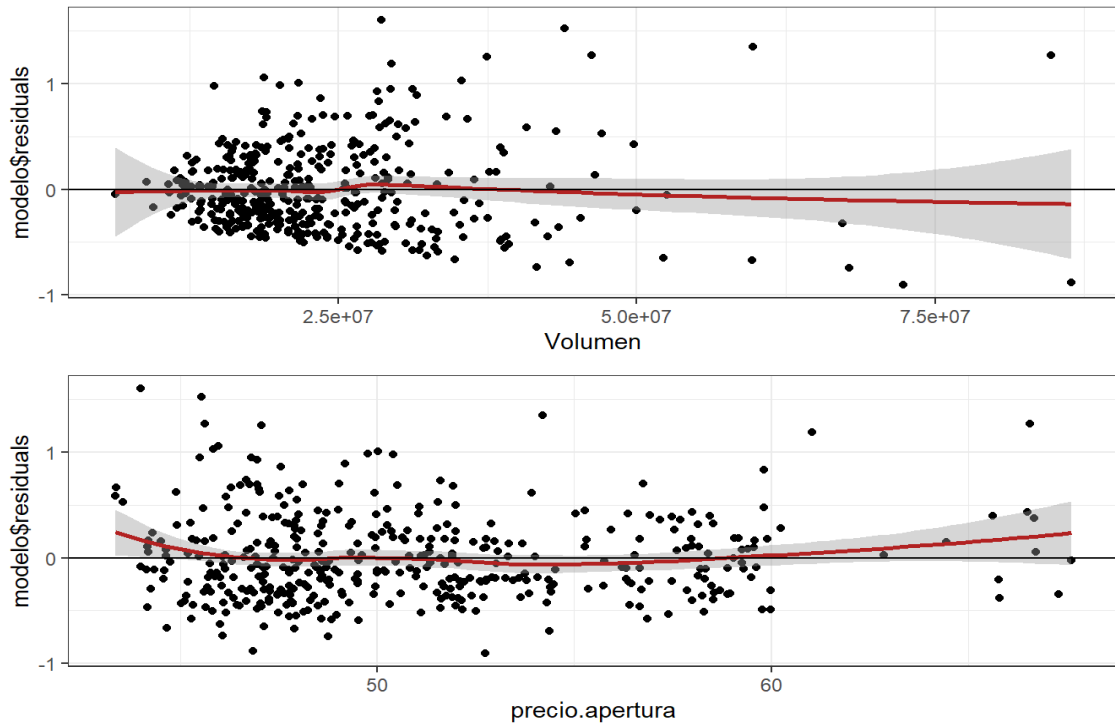


Figura 12. Linealidad de las variables precio apertura y volumen de las acciones Intel

Fuente: Gráfico obtenidos a través del software R (ver anexo 2).

De la figura 12 se observa el supuesto de linealidad de las variables independientes X 's, que evidentemente, al estar la línea dentro de los intervalos se observa linealidad para el precio de apertura y el volumen de ventas de las acciones de Intel 2019.

❖ Prueba gráfica.

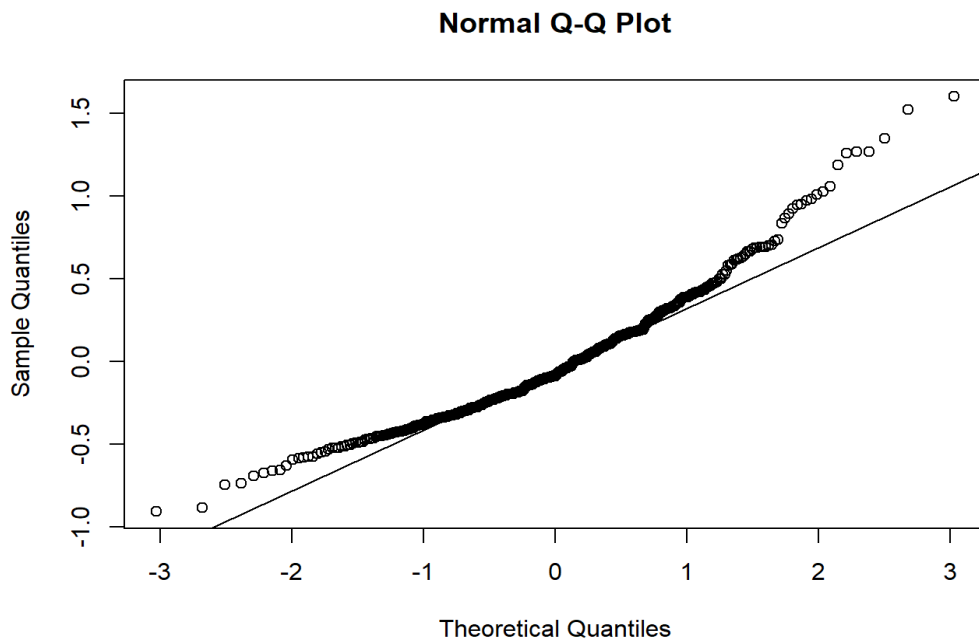


Figura 13. Gráfico de cuantiles para los errores del modelo

Fuente: Gráfico obtenidos a través del software R (ver anexo 2).

De la figura 13, se observa que los datos presentan normalidad de errores, en el apartado siguiente se realiza el supuesto de normalidad.

Prueba de hipótesis de normalidad (shapiro wilks)

1. Formulación de Hipótesis

H_0 : Existe normalidad en los errores (Hipótesis Nula)

H_a : No existe normalidad en los errores (Hipótesis Alternativa)

2. Nivel de significancia

$$\alpha = 5\% = 0.05$$

3. Prueba estadística



$$W = 0.95109, p\text{-value} = 2.096e-10$$

Valor calculado a partir de paquete estadístico R

4. Decisión

Si $p\text{-value} < \alpha$, se acepta H_a (altamente significativo)

Si $p\text{-value} > \alpha$, se acepta H_0 (no significativo)

5. Conclusión

Al ser $p(0.0000000002096) < \alpha(0.05)$, en este caso se acepta la hipótesis H_a , es decir, no existe normalidad en los errores, y se rechaza la hipótesis H_0

Prueba de homocedasticidad (test de Breusch Pagan)

1. Formulación de Hipótesis

H_0 : Existe homocedasticidad en los errores

H_a : No existe homocedasticidad en los errores.

2. Nivel de Significancia

$$\alpha = 5\% = 0.05$$

3. Prueba estadística

$$BP = 77.435, df = 2, p\text{-value} < 2.2e-16$$

4. Decisión

Si $p\text{-value} < \alpha$, se acepta H_a (altamente significativo)

Si $p\text{-value} > \alpha$, se acepta H_0 (no significativo)

5. Conclusión



Al ser $p(0.0000000000000000022) < \alpha(0.05)$, se acepta la hipótesis H_a ,
es decir, no existe homocedasticidad en las variables.

Prueba de autocorrelacion(Durwin Watson)

1. Planteamiento de Hipótesis

H_0 : Existe autocorrelación entre en las variables

H_1 : No existe autocorrelación entre las variables

2. Nivel de Significancia

$$\alpha = 5\% = 0.05$$

3. Prueba estadística

$$DW = 2.130209 \quad p\text{-value} < 0.176$$

4. Decisión

Si $p\text{-value} < \alpha$, se acepta H_a (altamente significativo)

Si $p\text{-value} > \alpha$, se acepta H_0 (no significativo)

5. Conclusión

$p(0.176) > \alpha(0.059)$, entonces, existe autocorrelación entre las
variables, es decir se acepta H_0 , y se rechaza H_a .

Tabla 3 . Resumen de los supuestos del primer modelo de regresión

Supuesto	Cumple
Linealidad	SI
Normalidad	NO
Homocedasticidad	NO
Autocorrelación	SI

Fuente: elaborado por el equipo de trabajo.

Debido a que no se cumple la normalidad en los errores y la heterocedasticidad entre las variables, el modelo no es confiable a pesar del R^2 es alto 0.9934 y el ANVA significativo.

Para corregir el modelo se optó por usar el método de eliminación de puntos Outliers. Debido a los valores atípicos o influyentes que se muestran con mucha preocupación en los diagramas de cajas anteriormente mostrados.

Identificación de valores atípicos

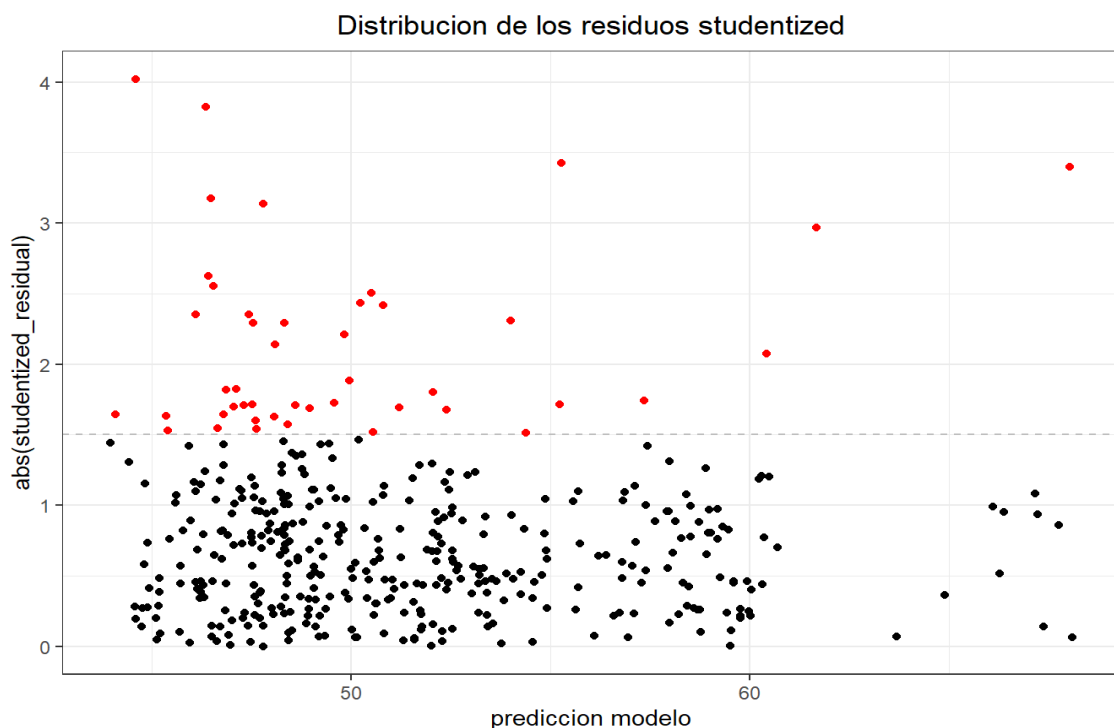


Figura 14. Gráfico de identificación de valores atípicos

Fuente: Resultados obtenidos a través del software R. (ver anexo 2 A).



El grafico 14, se puede apreciar muchos valores atípicos en un corte de 1.5. Estos valores son:

```
10 12 13 38 71 79 82 93 103 107 108 123 131 139 162 177 179 189 198 223 254 2
56 261 270 275 277 279 281 292 294 300 306 307 309 317 319 329 333 341 346 354 3
68 385 403 410
```

En esta parte del apartado, el software R nos muestra la posición de los puntos atípicos, los cuales afectan potencialmente a la estimación de los parámetros, por ende, se realizó la limpieza de datos, con el método de eliminación de puntos Outliers. (Ver anexo 2 A)

3.2.2. Obtención del segundo modelo de RLM.

Modelo de regresión lineal múltiple sin datos influyentes:

Coefficients:

	Estimate	Std.Error	t value	Pr(> t)
Intercepto	1.750e-02	1.754e-01	0.100	0.921
precio.apertura	1.003e+00	3.100e-03	323.529	< 2e-16 ***
volumen	1.199e-08	1.918e-09	6.252	1.14e-09 ***

Residual standard error: 0.2849 on 362 degrees of freedom

Multiple R-squared: 0.9968, Adjusted R-squared: 0.9967

F-statistic: 5.579e+04 on 2 and 362 DF, p-value: < 2.2e-16

Fuente: Resultados obtenidos a través del software R. (ver anexo 2 A).

R nos muestra los coeficientes de $\beta_0, \beta_1, \beta_2$ para este segundo modelo, luego de haber utilizado el método de eliminación de puntos Outlier.

Referencia

(***). Indica la alta Significancia.

El modelo de regresión lineal múltiple, queda definido como:

$$Y = 0.01750 + 1.00300X_1 + 0.0000001199X_2 \quad (9)$$

Las dos variables independientes precio de apertura y volumen resultan ser significativas $p < \alpha$ (0.05) excepto el intercepto. El modelo en su conjunto observando la prueba F es significativa y la bondad de ajuste es bastante alta $R^2 = 0.9968$.

Si observamos el error estándar residual este disminuye en comparación al modelo anterior (0.4072 frente a 0.2849 del modelo 2).

Comprobación del supuesto de normalidad

Observando la prueba de normalidad de Kolmogorov smirnov

1. Formulación de Hipótesis

H_0 : Existe normalidad en los errores

H_a : No existe normalidad en los errores

2. Prueba estadística

$D = 0.076623$, p-value = 0.02752

3. Decisión

Si p-value $< \alpha$, se acepta H_a (altamente significativo)

Si p-value $> \alpha$, se acepta H_0 (no significativo)

4. Conclusión

$p = 0.02752 < \alpha = 0.05$, en este caso se acepta la hipótesis H_a , es

decir, no existe normalidad en los errores, y se rechaza la hipótesis H_0

Segunda limpieza de datos atípicos

El modelo no está libre de normalidad en los errores.

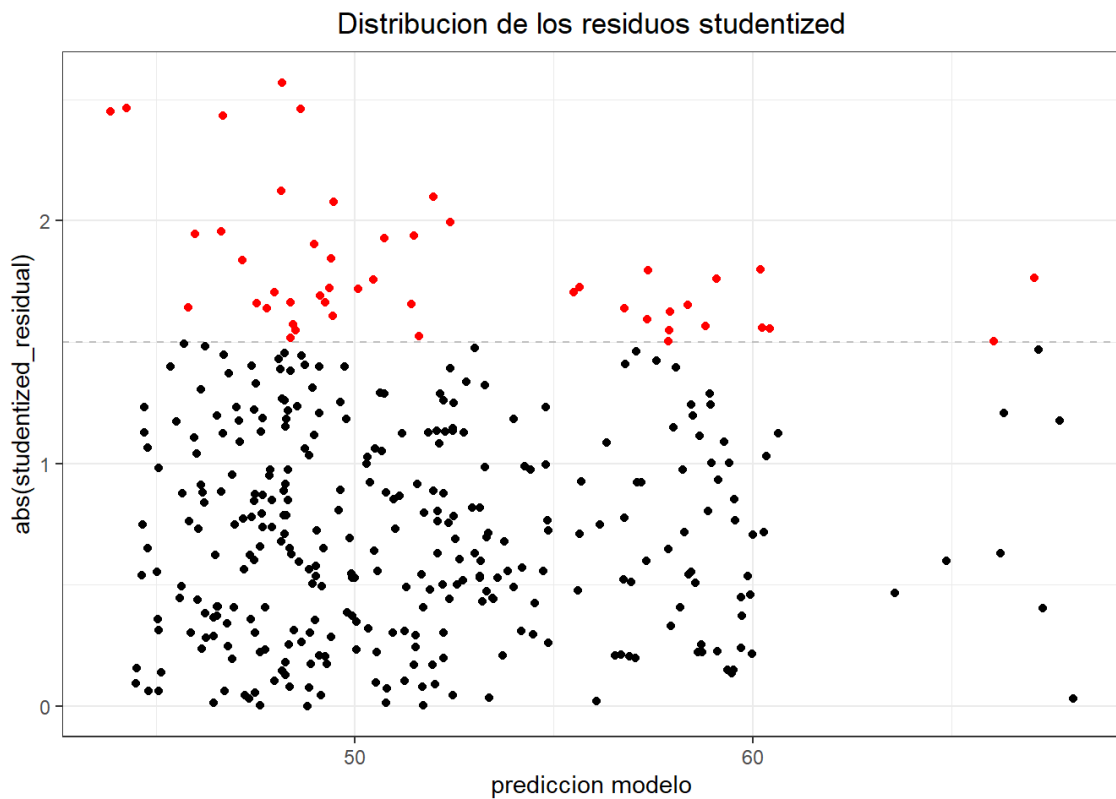


Figura 15. Gráfico de datos Estandarizados mayores a 1.5

Fuente: elaborado por el equipo de trabajo (anexo 2 B).

Se muestra los datos observados, son datos estandarizados mayores a 1.5. Estos valores son excluidos para poder obtener mejores resultados.

```
6 8 11 19 21 31 41 46 48 60 61 65 86 96 114 117 121 124 131 153 161 174 18
0 185 194 195 236 239 241 243 250 258 262 264 265 267 269 279 281 286 294 299
300 301 309 339 350 358
```

Salida de R: muestra los ítems de los valores influyentes (ver anexo 2).

Se identificaron valores influyentes en los ítems u observaciones, que se muestran, para ello se utilizó el método de eliminación de puntos Outlier y posterior a ello se calculó los parámetros para Regresión lineal múltiple.

3.2.3. Obtención del tercer modelo de RLM.

```
Coefficients:
              Estimate Std. Error  t value  Pr(>|t|)
Intercepto      2.856e-02  1.513e-01   0.189    0.85
precio.apertura 1.003e+00  2.685e-03  373.551  < 2e-16 ***
volumen         9.405e-09  1.681e-09   5.595    4.8e-08 ***

Residual standard error: 0.2257 on 314 degrees of freedom
Multiple R-squared:  0.9979, Adjusted R-squared:  0.9979
F-statistic: 7.448e+04 on 2 and 314 DF, p-value: < 2.2e-16
```

Fuente: Resultados obtenidos a través del software R. (ver anexo 2 B).

obtenidos a partir de la limpieza de datos.

El modelo de regresión lineal múltiple, queda definido como:

$$Y = 0.02856 + 1.003X_1 + 0.00000009405X_2 \quad (10)$$

Las dos variables independientes precio de apertura y volumen resultan ser significativas $p < \alpha$ (0.05) excepto el intercepto. El modelo en su conjunto observando la prueba F es significativa y la bondad de ajuste es bastante alta $R^2 = 0.9979$.

Si observamos el error estándar residual este disminuye en comparación al modelo anterior 0.2849 frente a 0.2257 del modelo 3.

Referencia

(***). Indica la alta Significancia.

INTERPRETANDO:

$$Y = 0.02856 + 1.003\text{PrecioApertura} + 0.00000009405\text{Volumen}$$

β_1 : Por cada unidad monetaria del precio de apertura en el cual la acción es abierta al mercado al mercado, el precio máximo de venta en un día, se ve **incrementada** en **1.003**.

β_2 : Por cada venta que se realiza el precio máximo de venta de las acciones de Intel, incrementa en **0.00000009405**

Comprobación de los supuestos:

Linealidad

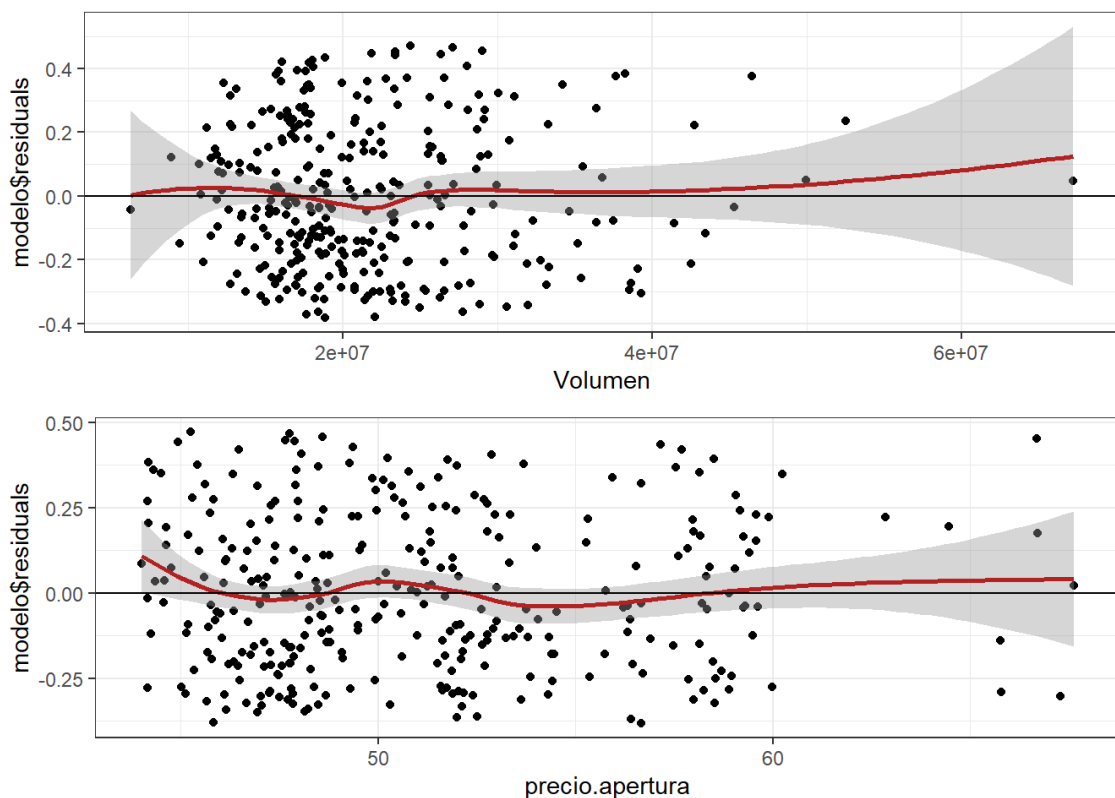


Figura 16. Gráfico de linealidad de las variables precio de apertura y volumen

Fuente: Resultados obtenidos a través del software R (Ver anexo 2B).



El grafico muestra que los datos son lineales.

Prueba de normalidad de kolmogorov smirnov

1. Formulación de Hipótesis

H_0 : Los residuos siguen una distribución normal

H_a : Los residuos no siguen una distribución normal

2. Nivel de Significancia

$$\alpha = 5\% = 0.05$$

3. Prueba Estadística

$$\diamond D = 0.068798, p\text{-value} = 0.09948$$

4. Decisión

Si $p\text{-value} < \alpha$, se acepta H_a (altamente significativo)

Si $p\text{-value} > \alpha$, se acepta H_0 (no significativo)

5. Conclusión

$p(0.09948) > \alpha(0.05)$ en este caso se acepta la hipótesis H_0 , se cumple que los residuos siguen una distribución normal, y se rechaza la hipótesis

H_a

Prueba de homocedasticidad (Breusch Pagan)

1. Formulación de Hipótesis

H_0 : Existe homocedasticidad en los errores

H_a : No existe homocedasticidad en los errores

2. Nivel de Significancia

$$\alpha = 5\% = 0.05$$



3. Prueba estadística

$$BP = 2.9952, df = 2, p\text{-value} = 0.2237$$

4. Decisión

Si $p\text{-value} < \alpha$, se acepta H_a (altamente significativo)

Si $p\text{-value} > \alpha$, se acepta H_0 (no significativo)

5. Conclusión

$p(0.2237) > \alpha(0.05)$, en este caso se acepta la hipótesis H_0 , es decir, existe homocedasticidad en los errores, y se rechaza la hipótesis H_a

Prueba de Autocorrelacion durwin watson

1. Formulación de Hipótesis

H_0 : Existe autocorrelación entre las variables

H_a : No existe autocorrelación entre las variables

2. Nivel de Significancia

$$\alpha = 5\% = 0.05$$

3. Prueba estadística

lag	Autocorrelation	D-W Statistic	p-value
1	0.03430749	1.925525	0.422

4. Decisión

Si $p\text{-value} < \alpha$, se acepta H_a (altamente significativo)

Si $p\text{-value} > \alpha$, se acepta H_0 (no significativo)



5. Conclusión

$p = 0.422 > \alpha = 0.05$, en este caso se acepta la hipótesis H_0 , es decir, existe Autocorrelación en los errores, y se rechaza la hipótesis H_a

Finalmente se observa la prueba VIF (Análisis de Inflación de Varianza)

volumen	precio.apertura
1.075775	1.075775

Se observa el valor en cuanto a la inflación de varianza ambos valores son menores que

$1 < VIF < 5$: La regresión puede verse afectada por cierta colinealidad.

Tabla 4. Resumen de los supuestos del tercer modelo de regresión.

Supuesto	Cumple
Linealidad	SI
Normalidad	SI
Homocedasticidad	SI
Autocorrelación	SI
Análisis de Inflación de Varianza	SI

Fuente: Elaborado por el equipo de trabajo.

Se observa el resumen de los supuestos del tercer modelo de la RLM, en este último modelo se cumplen todos los supuestos esperados.

Tabla 5. Resumen de errores de la Regresión lineal múltiple.

RESUMEN	Regresión Lineal Múltiple		
	Primer Modelo	Segundo Modelo	Tercer Modelo
R^2	0.9934	0.9968	0.9979
Residual Estándar Error	0.4072	0.2849	0.2257

Fuente: Elaborado Por el Equipo de Trabajo.

La tabla 6 muestra el resumen de la regresión lineal múltiple, utilizando el método de eliminación de puntos atípicos Outliers, se afirma que tercer modelo presenta menor Error estándar residual de 0.2257, y un valor Alto R^2 0.9979.

3.3. OBTENCIÓN DEL ARBOL DE REGRESIÓN.

Para obtener el árbol de regresión se particiono la data en datos de entrenamiento y validación, Particionando los datos en 80% para entrenamiento y 20% para validación, siendo el total de la muestra 410 registros ,328 para entrenamiento y 82 para validación.



Figura 17. Numero de hojas entrenamiento y validación

Fuente: figura obtenida a través del software R.

De la figura 16, se determinó el número de nodos del Árbol de regresión, se observa la variación de nodos en 1, 2, 3 y 4, los nodos siguientes no presentan mucha variación en cuanto al error, sumado a esto el AR al predecir nuevos casos no estima con el mismo índice de acierto, por ello se determina el desarrollo de 4 nodos de profundidad.

3.3.1. Árbol regresión utilizando el tipo uniforme:

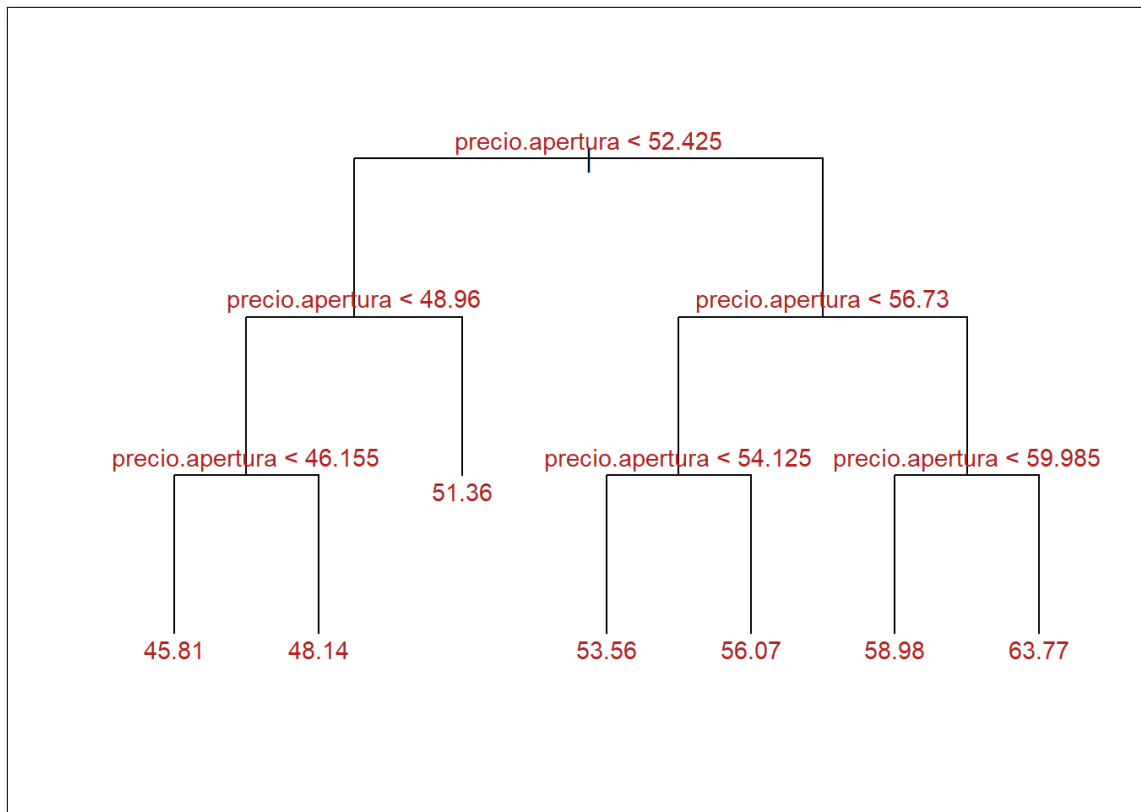


Figura 18. Árbol de Regresión utilizando el tipo Uniforme para predecir.

Fuente: Resultados obtenidos a través del software R.

Se muestra los resultados del Árbol de regresión utilizando el algoritmo de tipo uniforme, con las variables más influyentes, para este estudio se utilizaron el Precio máximo, y el Precio de Apertura de las acciones de Intel.



Interpretación:

Intel alcanza el siguiente valor máximo de acciones:

- ❖ \$63.77 cuando en precio de apertura es mayor a \$59.985, mayor a \$56.73 y mayor a \$52.425.
- ❖ \$58.98 cuando en precio de apertura es menor a \$59.985, mayor a \$56.73 y mayor a \$52.425
- ❖ \$56.07 cuando en precio de apertura es mayor a \$54.125, menor a \$56.73 y mayor a \$52.425
- ❖ \$53.56 cuando en precio de apertura es menor a \$54.125, menor a \$56.73 y mayor a \$52.425
- ❖ \$51.36 cuando en precio de apertura es mayor a \$48.96 y menor a \$52.425
- ❖ \$48.14 cuando en precio de apertura es mayor a \$46.155, menor a \$48.96 y menor a \$52.425
- ❖ \$45.81 cuando en precio de apertura es menor a \$46.155, menor a \$48.96 y menor a \$52.425



Otra forma de representar, el AR

```
## node), split, n, deviance, yval
##      * denotes terminal node
##
## 1) root 287 5911.000 51.18
##    2) precio.apertura < 52.425 204 1085.000 48.79
##      4) precio.apertura < 48.96 130 243.700 47.32
##        8) precio.apertura < 46.155 46 34.730 45.81 *
##        9) precio.apertura > 46.155 84 47.040 48.14 *
##      5) precio.apertura > 48.96 74 69.970 51.36 *
##    3) precio.apertura > 52.425 83 774.300 57.07
##      6) precio.apertura < 56.73 42 83.880 54.64
##        12) precio.apertura < 54.125 24 6.405 53.56 *
##        13) precio.apertura > 54.125 18 12.740 56.07 *
##      7) precio.apertura > 56.73 41 186.500 59.57
##        14) precio.apertura < 59.985 36 35.320 58.98 *
##        15) precio.apertura > 59.985 5 50.370 63.77 *
```

(*) representa lo valores predichos.

Para la validación del árbol de regresión haciendo uso del 20% de la data se obtiene lo siguiente:

```
## [1] "Error de test (mse) del arbol de regresion tras podado: 1.448015"
```

Se realizo la validación con el 20% de la data obteniendo un MSE de 1.448015, teniendo este resultado procedemos a podar el AR para mejorar y obtener un MSE mucho mejor del que se tiene.

3.3.2. Árbol de regresión algoritmo uniforme (Poda)

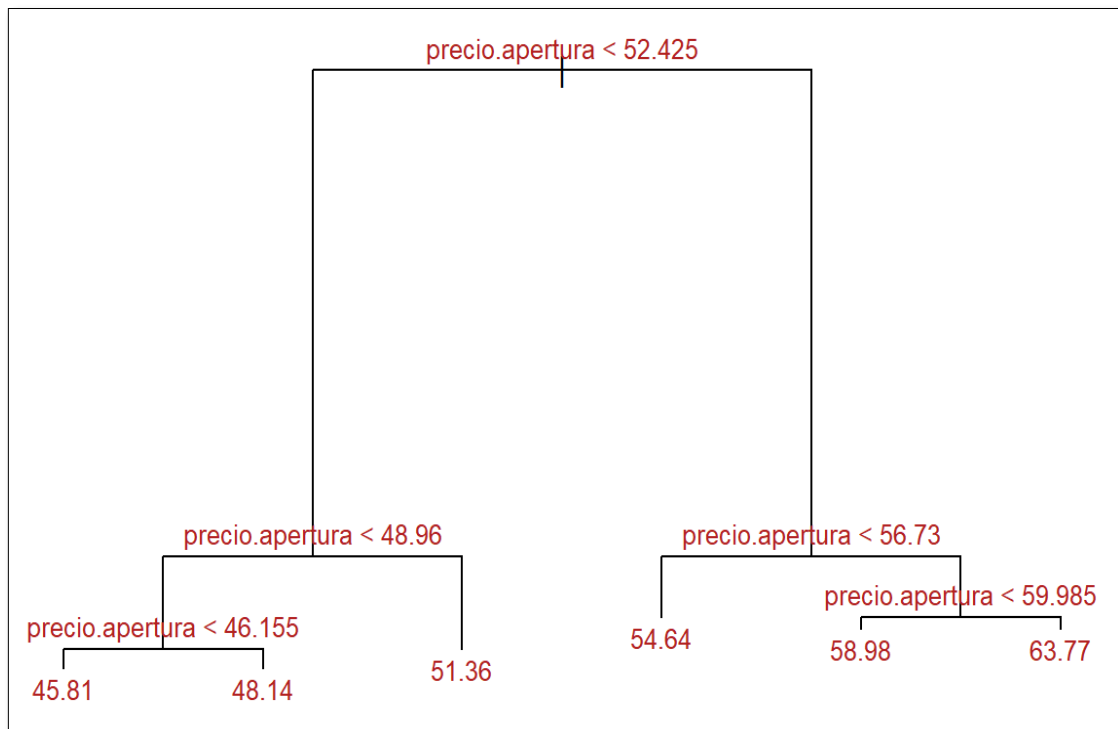


Figura 19. Representación del Árbol de Regresión podado con el algoritmo uniforme.

Fuente: Resultados obtenidos a través del software R.

De la figura 20 se observa el podado del AR, se muestra 6 nodos terminales, o también llamados valores predichos, evidentemente se redujo, el nodo precio de apertura < a \$54,64.

Evaluando el árbol Regresión con el 20% de la data.

```
## [1] "Error de test (mse) del arbol de regresion tras podado: 1.61"
```

Al validar el Árbol de Regresión con el 20% de la data se obtuvo el error MSE de 1.61, se realiza el podado del árbol para obtener menor error, pero en este caso paso lo contrario, debido a que se trabajó con un número reducido variables.

Tabla 6. Comparación de errores del Árbol de regresión

COMPARACIÓN	Árbol de Regresión	
	Sin podar	Podado
MSE	1.448015 \$	1.61 \$

Fuente: Elaborado por el equipo de trabajo

Se observa, que sin podar el AR se obtiene 1.448015\$ < al podado 1.61 \$, se concluye no podar el árbol, puesto que esta presenta menos error.

3.4. RESUMEN DE LOS ALGORITMOS

Tabla 7. Resumen de errores para la Regresión lineal y Árbol de regresión

ALGORITMO	REGRESIÓN LINEAL MÚLTIPLE	ÁRBOL DE REGRESIÓN
	Tercer Modelo	Validación sin podado
R^2	0.9934	_____
Error Estándar /MSE	\$ 0.2257	\$ 1.448015

Fuente: elaborado por el equipo de trabajo

Se muestra el resumen de los errores MSE y R^2 de Regresión lineal múltiple con la eliminación de puntos atípicos Outlier, obteniendo de esta forma un R^2 muy bueno 0.9979 lo cual indica que el 0.99% de los datos se ajustan al modelo y el Error estándar residual con \$ 0.2257, seguidamente se muestra el algoritmo Árbol de Regresión, para la validación sin podado se encontró un MSE de \$1.4480; y finalmente se concluye que la Regresión lineal múltiple tiene menor error frente al Árbol de regresión.



Discusión de los resultados

Lizares (2017) comparo la Regresión logística binaria y Árboles de clasificación (CHAID) para evaluar el rendimiento académico. Concluye que, Según la evaluación de la clasificación de los modelos optamos por la Técnica de Árboles de clasificación, siendo la más óptima por tener mayor Sensibilidad=77,6% AUC=90,1%, Gini =80,2% y Kappa=0,589. A diferencia de esta tesis, en la que se empleó el Árbol de regresión por la presencia de variables cuantitativas, se encuentro mayor MSE \$1.448015 comparado al Error Estándar Residual de la RLM.

También Espinosa (2015) quien concluye que realizando un análisis técnico de los datos históricos del valor de las acciones durante un plazo determinado, es posible predecir el cambio de precio de una acción, con la finalidad de ayudar al inversor al momento de tomar la decisión de compraventa trading de sus acciones, utilizando para esto regresión múltiple, lo que concuerdan con los resultados del presente trabajo de investigación al tener la regresión lineal múltiple un Error Estándar de 0.2257, se confirma que el modelo de Regresión múltiple es buen candidato al momento de enfrentarse a un problema de predicción de variables continuas.

Confrontando con Sepúlveda & Morales (2013) que, en su estudio comparo árbol regresión CART y Regresión Lineal, encontró que, el error predictivo de la regresión lineal siempre es menor que el del CART. Aconteció algo similar en esta tesis, en la que adicionalmente se emplearon técnicas estadísticas como la eliminación de puntos Outlier para la Regresión lineal múltiple y la poda en el árbol de regresión, sin embargo, se concluyó que el modelo de regresión lineal múltiple tiene menor error frente el árbol de regresión para predecir el precio máximo en funcia del precio de apertura y el volumen de ventas de Intel.



V. CONCLUSIONES

PRIMERA: Se logró construir el modelo de regresión lineal múltiple para poder predecir el precio máximo del valor de las acciones en función al precio de apertura y el volumen de ventas de Intel, con la técnica de eliminación de puntos atípicos Outliers, queda definida por la siguiente ecuación. $Y = 0.02856 + 1.003X_1 + 0.000000009405X_2$, Alcanzando una prueba F significativa y la bondad de ajuste es bastante alta $R^2 0.9934$ y un error estándar residual de \$ 0.2257.

SEGUNDA: Se logró obtener el árbol de regresión para predecir el precio máximo de las acciones en función al precio de apertura y el volumen de ventas de las acciones de Intel, concluye que la variable para explicar el precio máximo de acciones de Intel por día es el precio de apertura, eliminando la variable volumen de ventas. Se optó por el árbol de regresión sin podado puesto que el error MSE fue de \$ 1.448015.

TERCERA: Y finalmente se concluye que el mejor modelo para predecir el precio máximo de acciones de Intel es el modelo de regresión lineal Múltiple con eliminación de Outliers, debido a que este presenta menor error frente a un árbol de regresión. Este modelo ayudara a reducir la incertidumbre en los Accionista o Inversionistas de la empresa Intel.



VI. RECOMENDACIONES

PRIMERA: A futuros investigadores tener en cuenta este trabajo de investigación seguir innovando y hacer uso de las técnicas estadísticas, así como Regresión lineal múltiple y otros, así como en este trabajo de investigación se puede ir complementando con otras técnicas aparte de la eliminación de puntos atípicos Outlier, realizar el tratamiento de datos. Los resultados encontrados en este trabajo pueden aportar significativamente al plantear nuevas propuestas ya sean académicas o metodológicas.

SEGUNDA: Tener en cuenta los árboles de regresión como una alternativa porque tiene una gran capacidad predictiva y puede aprender patrones no lineales que los modelos de regresión lineal múltiple no lo hacen.

TERCERA: Utilizar el modelo de regresión lineal múltiple obtenido en esta tesis para realizar estimaciones de esta esta forma obtener decisiones acertadas.



VII. REFERENCIAS BIBLIOGRÁFICAS

- Alvarez, M. L. (2016). Predicción de afinidad de unión de ligandos en proteínas vía Aprendizaje Automático. Universidad De Buenos Aires.*
- Arenas, G. A. D. L. (2011). Modelos de regresión con problemas de autocorrelación (Doctoral dissertation, CENTRO DE INVESTIGACIÓN EN MATEMÁTICAS).*
- Ayala Jacinto, A. R., & Samillán Vásquez, C. J. (2019). Metodología de Box Jenkins vs Redes Neuronales Artificiales para construir un modelo de pronóstico del precio de compra de cierre mensual de las acciones del Banco de Crédito del Perú en La Bolsa de Valores de Lima, abril de 2005 hasta febrero de 2018.*
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, 1287-1294.*
- Camborda Zamudio, M. G. (2014). Aplicación de árboles de decisión para la predicción del rendimiento académico de los estudiantes de los primeros ciclos de la carrera de ingeniería civil de la universidad continental. Universidad Nacional Del Centro Del Perú.*
- Candia Oviedo, D. I. (2019). Predicción del rendimiento académico automático, estudiantes de la unsaac a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automatico. Universidad Nacional San Antonio Abad Del Cusco.*
- Dupouy Berrios, C. (2014). Aplicación de árboles de decisión para la estimación del escenario económico y la estimación de movimiento la tasa de interés en chile. Universidad De Chile.*



- Eainversores. (3 de julio del 2020). Como interpretar el volumen. Recuperado de <https://eainversores.com/wp-content/uploads/2017/10/Como-interpretar-el-Volumen.pdf>*
- Espinosa Muñoz, J. I. (2015). Una aproximación a la predicción del valor de acciones en la bolsa de Valores aplicando técnicas de Data Mining.*
- Espitia, c. G. G. El modelo de regresión lineal múltiple con stata♦.*
- Garcia Sandoval, J. P. (2010). Predictibilidad en los retornos de la acción de Ecopetrol: evidencia a través del efecto día. Universidad Pontificia Bolivariana.*
- Giménez Fernández, R., & Zamorano Cid, P. (2014). Modelos Predictivos de Índices Bursátiles Relevantes para la Economía Chilena. Universidad De Chile.*
- Gordón Mendoza, R., & Camargo-Buitrago, I. (2015). Selección de estadísticos para la estimación de la precisión experimental en ensayos de maíz. Agronomía Mesoamericana, 26(1), 56-63.*
- Hernandez Sampieri, R. (2014). Metodología de la Investigación (6a ed.). Mexico: McGraw Hill*
- Heredia, S. R., Bruno, Cecilia, & Balzarini, Monica (2010). Identificación de relaciones entre rendimientos y variables ambientales vía árboles de clasificación y regresión (CART). Interciencia, 35(12), 876-882.*
- Kleinbaum D., Kupper L. (1998). Applied Regression Analysis and other Multivariable Methods. 3rd. Ed. Massachusetts: Duxbury Press.*
- Leon Atiquipa, H. E. (2018). Desarrollo De Un Modelo Algorítmico Basado En Árboles De Decisión Para La Predicción De La Permanencia De Un Paciente En Un Proceso Psicoterapéutico. Pontifice Universidad Catolica Del Perú.*



- Lizares Castillo, M. (2017). Comparación de modelos de clasificación: regresión logística y árboles de clasificación para evaluar el rendimiento académico.*
- Pérez, E. R., & Medrano, L. A. (2010). Análisis factorial exploratorio: bases conceptuales y metodológicas. Revista Argentina de Ciencias del Comportamiento (RACC), 2(1), 58-66.*
- Peña, A. M. (2011). Análisis de la rentabilidad-riesgo de acciones que cotizan en la bolsa de valores de Colombia para el periodo 2002-2008. Universidad del valle.*
- Rizola Margolis, P. E. (2017). Rentabilidad y riesgo en el mercado bursátil mexicano. Universidad Politecnica De Cataluña.*
- Sampieri, R. H., Collado, C. F., Lucio, P. B., Valencia, S. M., & Torres, C. P. M. (1998). Metodología de la investigación (Vol. 1, pp. 233-426). México, DF: Mcgraw-hill.*
- Sepúlveda, J. F. D., & Morales, J. C. C. (2013). Comparación entre árboles de regresión CART y regresión lineal. Comunicaciones en Estadística, 6(2), 175-195.*
- Tobergte, D. R., & Curtis, S. (2013). Introducción a la econometría: Un enfoque moderno. Journal of Chemical Information and Modeling, 53.*
- Torres, M., Salazar, F. G., & Paz, K. (2019). Métodos de recolección de datos para una investigación.*
- Uriel, E. (2013). Regresión lineal múltiple: estimación y propiedades. Universidad de Valencia Versión, 09-2013.*
- Vallejo, P. M. (2012). Correlación y regresión, simple y múltiple. Universidad Pontificia Comillas, Madrid. Recuperado de <http://web.upcomillas.es/personal/peter/investigacion/Regresion.pdf>.*



Yamamoto, M. (2017). Análisis de los datos: detección de outliers. Contacto Científico, 7(1).

Ramírez, A. M. A. (2013). Métodos utilizados para el pronóstico de demanda de energía eléctrica en sistemas de distribución (Doctoral dissertation, Universidad Tecnológica de Pereira. Facultad de Ingenierías Eléctrica, Electrónica, Física y Ciencias de la Computación. Ingeniería Eléctrica.).

<https://www.maximaformacion.es/blog-dat/que-es-la-correlacion-estadistica-y-como-interpretarla/>



ANEXOS



ANEXO 1. CÓDIGO UTILIZADO EN R

library(caret)//paquete que divide los datos en 80% y 20%

library(ggplot2)//paquete para realizar gráficos

library(lmtest)//paquete para la prueba de homocedasticidad

library(rpart)//paquete para generar el árbol de decisión

library(rpart.plot)//paquete para graficar el árbol de decisión

library(MLmetrics)/paquete para calcular el RMSE

library(car)//paquete para gráficos

library(GGally)//paquete para gráficos

.....

#lectura de datos

.....

```
datos <- read.csv("C:/Users/User/Downloads/datos.csv")
```

```
names(datos)[names(datos)=='Open']<-'precio.apertura'
```

```
names(datos)[names(datos)=='High']<-'precio.maximo'
```

#distribucion de las variables

```
ggplot(datos, aes(x = precio.maximo)) +
```

```
  geom_histogram(aes(y = ..density..), # the histogram will display "density" on its y-axis
```

```
    binwidth = .5, colour = "blue", fill = "white") +
```

```
  geom_density(alpha = .2, fill="#0072f3") +
```

```
  geom_vline(aes(xintercept = mean(precio.maximo, na.rm = T)),
```

```
    colour = "red", linetype = "longdash", size = .8)
```

```
ggplot(datos, aes(x = precio.apertura)) +
```

```
  geom_histogram(aes(y = ..density..), # the histogram will display "density" on its y-axis
```

```
    binwidth = .5, colour = "blue", fill = "white") +
```

```
  geom_density(alpha = .2, fill="#0072f3") +
```

```
  geom_vline(aes(xintercept = mean(precio.apertura, na.rm = T)),
```

```
    colour = "red", linetype = "longdash", size = .8)
```

```
ggplot(datos, aes(x = Volume)) +
```



```
geom_histogram(aes(y = ..density..)
  , colour = "blue", fill = "white") +
geom_density(alpha = .2, fill="#0072f3") +
geom_vline(aes(xintercept = mean(Volume, na.rm = T)),
  colour = "red", linetype = "longdash", size = .8)
shapiro.test(modelo$residuals)
require(corrplot)
corrplot(cor(select(datos, habitantes, asesinatos, universitarios, heladas)), method =
"number")
.....
#boxplots
.....
ggplot(datos, aes(y = precio.maximo)) + geom_boxplot(fill="#56B4E9", outlier.shape =
22,colour="black")
ggplot(datos, aes(y = precio.apertura)) + geom_boxplot(fill="#56B4E9", outlier.shape =
22,colour="black")
ggplot(datos, aes(y = Volume)) + geom_boxplot(fill="#56B4E9", outlier.shape =
22,colour="black")
.....
#partición 80% para train y 20% para test
.....
set.seed(1212)
indice <- createDataPartition(datos$precio.maximo,p=.8,list = FALSE,times=1)
train<-datos[indice,2:4]
test<-datos[-indice,2:4]
.....
#modelo de regresion linea multiple
.....
mod<-lm(precio.maximo~.,data=train)
summary(mod)
mod <- fortify(mod)
```



```
ggplot(mod,aes(x=precio.apertura,y=.resid))+geom_point()+geom_smooth(colour="red")
)+geom_hline(yintercept = 0)

ggplot(mod,aes(x=Volume,y=.resid))+geom_point()+geom_smooth(colour="red")+geom
_hline(yintercept = 0)

ggplot(mod,aes(sample=.resid))+stat_qq()+stat_qq_line()

ggplot(mod,aes(x=.fitted,y=.resid))+geom_point()+geom_smooth(colour="red")+geom
_hline(yintercept = 0)

ks.test(mod$.resid,"pnorm", mean(mod$.resid), sd(mod$.resid))

mod<-lm(precio.maximo~.,data=train)

bptest(mod)

vif(mod)

.....

#arbol de regresión

.....

library(MASS)

# data("INTEL")

Boston <- read.csv("DATOS-INTEL.csv",head=T, sep=",")

head(Boston)

str(Boston)

# exportando datos Boston a formato texto

# write.csv(Boston, file="Boston.csv", row.names = F)

# Ajuste del modelo

library(tree)

set.seed(1) # para reproducir los mismos resultados

# Para dividir la data en dos grupos cada una con 50% de datos

# train <- sample(1:nrow(Boston), size = nrow(Boston)/2)

# creando data train (80%) y test 20%

train<- sample(1:nrow(Boston), 0.7 * nrow(Boston))

test<- setdiff(1:nrow(Boston), train)
```



```
# variable respuesta medv

arbol_regresion <- tree(precio.maximo ~ ., data = Boston, subset = train,
  split = "deviance")

summary(arbol_regresion)

library(ggplot2)
library(magrittr)
library(ggpubr)

resultados_cv <- data.frame(n_nodos = cv_arbol$size, deviance = cv_arbol$dev,
  alpha = cv_arbol$k)

p1 <- ggplot(data = resultados_cv, aes(x = n_nodos, y = deviance)) +
  geom_line() +
  geom_point() +
  labs(title = "Error vs tamaño del árbol") + theme_bw()

p2 <- ggplot(data = resultados_cv, aes(x = alpha, y = deviance)) +
  geom_line() +
  geom_point() +
  labs(title = "Error vs hiperparametro alpha") + theme_bw()

ggarrange(p1, p2)

#####

#uniform

plot(x = arbol_regresion, type = "uniform")

text(x = arbol_regresion, splits = TRUE, pretty = 0, cex = 0.9, col = "firebrick")

arbol_regresion

# Evaluacion del modelo

predicciones <- predict(arbol_regresion, newdata = Boston[test,])

test_mse <- mean((predicciones - Boston[test, "precio.maximo"])^2)

paste("Error de test (mse) del árbol de regresion tras podado:", round(test_mse,6))

# Podado del árbol (pruning)

set.seed(3) #semilla del generador de numeros aleatorios de R
```



```
cv_arbol <- cv.tree(arbol_regresion, K = 6)
```

```
cv_arbol
```

ANEXOS 2. BASE DE DATOS OBTENIDO DE NASDAQ

OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
1	\$67.40	\$67.30	17408040	36	\$58.57	\$58.21	19135290
2	\$67.60	\$66.69	23401420	37	\$58.34	\$57.55	22795850
3	\$66.06	\$65.77	20970750	38	\$58.05	\$56.75	27896160
4	\$65.04	\$64.46	16654560	39	\$57.17	\$56.91	20051870
5	\$65.98	\$65.80	25268430	40	\$56.83	\$56.71	16157980
6	\$66.50	\$65.64	18522360	41	\$56.96	\$56.54	14467820
7	\$68.06	\$67.63	18076690	42	\$56.99	\$56.66	16477410
8	\$67.58	\$66.50	30915390	43	\$56.41	\$56.41	17675580
9	\$67.42	\$66.72	30764000	44	\$56.66	\$56.66	18882730
10	\$69.29	\$66.57	84711040	45	\$56.86	\$56.86	26613650
11	\$63.71	\$62.86	42778740	46	\$58.69	\$58.55	19166770
12	\$62.85	\$61.04	29475120	47	\$58.30	\$58.15	9481129
13	\$61.25	\$59.81	28358470	48	\$58.59	\$58.53	18185510
14	\$60	\$59.98	21803450	49	\$59.13	\$58.95	22381500
15	\$59.84	\$59.26	21365530	50	\$58.84	\$58	15012420
16	\$59.65	\$59.30	18498800	51	\$58.65	\$58.34	15690160
17	\$59.74	\$59.49	17092450	52	\$58.30	\$57.49	16139030
18	\$59.78	\$59.17	16455830	53	\$58.36	\$58.26	18559090
19	\$60.08	\$59.57	15201930	54	\$58.65	\$58.48	17062830
20	\$59.51	\$59.05	18818490	55	\$58.48	\$57.70	16101280
21	\$59.32	\$58.89	23138270	56	\$58.70	\$58.17	16112990
22	\$59.80	\$59.78	21960520	57	\$58.04	\$57.61	12141220
23	\$60.20	\$59.59	17759140	58	\$57.97	\$57.87	15409830
24	\$60.70	\$59.81	15298580	59	\$58.80	\$58.40	11954040
25	\$60.97	\$60.24	17713400	60	\$58.50	\$57.97	11236370
26	\$59.93	\$59.62	14435830	61	\$58.30	\$57.85	11890270
27	\$60.05	\$59.99	12751110	62	\$58.55	\$58	16962850
28	\$60.48	\$59.91	14086150	63	\$57.71	\$57.49	17721000
29	\$59.85	\$59.41	11486640	64	\$58.35	\$57.47	20865350
30	\$59.48	\$59.25	6313179	65	\$57.80	\$56.93	19067980
31	\$59.78	\$59.06	23593190	66	\$56.65	\$55.94	17953470
32	\$59.05	\$58.32	49970820	67	\$56.60	\$56.33	19249680
33	\$57.99	\$57.17	18869000	68	\$56.63	\$56.32	15248000
34	\$57.46	\$57.39	28273060	69	\$57.24	\$56.37	18155770
35	\$58.10	\$58	21656380	70	\$56.81	\$56.38	32307510



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
71	\$56.61	\$54.19	59767680	106	\$50.99	\$50.25	17067940
72	\$52.43	\$52.09	25654060	107	\$51.17	\$49.93	30023120
73	\$51.96	\$51.37	15961600	108	\$48.95	\$47.56	23472110
74	\$52.32	\$51.90	13347320	109	\$47.22	\$47.12	15010270
75	\$52.17	\$51.52	13120650	110	\$47.79	\$47.24	16923150
76	\$51.82	\$51.75	16975160	111	\$47.22	\$46.46	17806110
77	\$52.84	\$52.64	16918320	112	\$45.91	\$45.70	14895350
78	\$52.56	\$52.35	15173070	113	\$46.12	\$45.87	16927570
79	\$53.07	\$51.93	18970760	114	\$45.82	\$45.82	22086630
80	\$51.95	\$51.90	16425530	115	\$46.63	\$46.35	32817080
81	\$52.79	\$52	24167110	116	\$47.43	\$47.28	19785890
82	\$51.78	\$50.41	14611660	117	\$47.29	\$47.11	15915800
83	\$50.79	\$50.47	12189960	118	\$47.12	\$47.03	23115320
84	\$50.48	\$50.16	17884430	119	\$47.60	\$47.46	21404920
85	\$51.23	\$50.70	12722050	120	\$46.68	\$46.34	23335690
86	\$50.97	\$50.35	12748070	121	\$46.18	\$46.10	21992680
87	\$50.07	\$49.50	17235570	122	\$46.63	\$46.06	25650220
88	\$50.37	\$50.31	21437000	123	\$47.05	\$45.49	29374660
89	\$52.24	\$51.97	18487340	124	\$46.07	\$45.76	18494300
90	\$51.60	\$51.19	14059200	125	\$47	\$46.94	24975460
91	\$51.22	\$50.63	14846410	126	\$47.37	\$46.16	30643740
92	\$51.15	\$50.98	11482110	127	\$46.88	\$46.30	29440390
93	\$51.21	\$49.75	20120840	128	\$47.56	\$47.15	26119560
94	\$51.18	\$51.10	28472930	129	\$47.96	\$47.76	38936650
95	\$51.23	\$50.83	21574150	130	\$49.36	\$49.09	27881620
96	\$51.88	\$51.60	38632500	131	\$51.89	\$50.52	34020780
97	\$51.99	\$51.70	11849920	132	\$51.72	\$51.51	24169750
98	\$52.13	\$52.13	15029860	133	\$52.31	\$51.71	16487690
99	\$52.14	\$52.05	19888220	134	\$52.56	\$51.53	21897460
100	\$52.29	\$51.90	13354590	135	\$53.50	\$52.69	36401550
101	\$53	\$52.76	18010810	136	\$52.60	\$52.51	27765430
102	\$53.33	\$53	23308710	137	\$52.97	\$51.98	20896370
103	\$52.79	\$51.60	18968940	138	\$52	\$52	18413520
104	\$51.84	\$51.33	18532030	139	\$51.52	\$50.01	21686530
105	\$52.03	\$51.06	20749660	140	\$50.66	\$49.96	29141090



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
141	\$49.99	\$49.28	15679640	176	\$44.53	\$43.35	28450340
142	\$49.70	\$49.50	11928610	177	\$44.69	\$44.66	34779860
143	\$50	\$49.91	17131370	178	\$45.05	\$44.59	27168070
144	\$50.25	\$50	14345080	179	\$44.74	\$43.37	35783930
145	\$49.93	\$49	20148730	180	\$44.46	\$44.15	16932520
146	\$48.78	\$48.56	16615080	181	\$44.78	\$44.34	25543370
147	\$48.61	\$47.92	16032270	182	\$44.52	\$44	28674970
148	\$47.80	\$47.36	17853190	183	\$45.73	\$45.18	22500360
149	\$47.95	\$47.62	16564290	184	\$45.95	\$45.47	19954650
150	\$48.18	\$48.03	14339230	185	\$45.75	\$44.92	23407890
151	\$48.56	\$48.18	10739100	186	\$45.49	\$45.18	24706460
152	\$48.22	\$47.91	15384690	187	\$45.64	\$45.33	39091930
153	\$49.45	\$49.29	27505450	188	\$46.80	\$46.44	42522780
154	\$48.20	\$47.77	26646180	189	\$48.29	\$47.90	59642160
155	\$48.42	\$48.15	23062950	190	\$50.79	\$50.20	36812430
156	\$48.48	\$47.65	21886310	191	\$50.88	\$50.60	29713380
157	\$47.68	\$47.56	16665580	192	\$51.28	\$50.25	26245240
158	\$47.85	\$47.40	14479940	193	\$51.82	\$51.08	25656640
159	\$47.73	\$46.95	31096250	194	\$51.33	\$50.80	22655400
160	\$48.05	\$47.87	20126780	195	\$51.60	\$51.10	20467580
161	\$47.58	\$47.48	17436380	196	\$51.43	\$51	25617960
162	\$48	\$46.75	25732290	197	\$52.43	\$52.17	38546980
163	\$46.53	\$46.14	12643340	198	\$53.09	\$52.74	72319400
164	\$46.42	\$45.83	15381080	199	\$58.38	\$58.30	27494710
165	\$46.96	\$46.60	12267630	200	\$59.29	\$58.52	17603860
166	\$46.67	\$46.67	15900570	201	\$58.98	\$58.89	17016380
167	\$47.32	\$47.26	18415500	202	\$58.83	\$58.15	12313140
168	\$47.19	\$46.23	20806080	203	\$58.88	\$58.71	20014510
169	\$46.12	\$45.25	24370250	204	\$59.59	\$58.44	38572210
170	\$45.17	\$44.75	19076170	205	\$57.47	\$56.66	30057100
171	\$45.18	\$45.11	20802010	206	\$56.56	\$56.46	10986910
172	\$44.86	\$44.16	28094920	207	\$56.49	\$56.21	12639630
173	\$44.59	\$44.25	31155680	208	\$56.07	\$55.76	10807340
174	\$44.79	\$44.18	25532530	209	\$55.86	\$55.32	12841760
175	\$45.03	\$44.30	21618860	210	\$55.43	\$55.35	13204510



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
211	\$55.74	\$55.28	11751750	246	\$51.67	\$51.25	22512380
212	\$55.95	\$55.75	18919000	247	\$51.04	\$50.41	17530450
213	\$56.14	\$55.27	16421070	248	\$50.81	\$50.13	17501310
214	\$55.99	\$55.03	24439760	249	\$50.14	\$49.36	18039220
215	\$54.80	\$54.53	13507980	250	\$49.32	\$49.02	17998810
216	\$54.57	\$54.34	17216940	251	\$49.02	\$48.77	18848690
217	\$54.04	\$53.80	18568690	252	\$49.85	\$49.50	27807690
218	\$53.44	\$53.25	13479770	253	\$50.64	\$49.95	21642300
219	\$53.86	\$53.59	19185200	254	\$50.72	\$49.18	31577230
220	\$53.77	\$53.32	17709670	255	\$49.24	\$48.62	20851710
221	\$53.55	\$52.97	16638820	256	\$49.05	\$47.06	37424720
222	\$54.62	\$54.40	21329170	257	\$47.18	\$46.70	41457610
223	\$54.99	\$53.92	18715860	258	\$47.68	\$46.68	26746480
224	\$54.40	\$54.06	23330720	259	\$46.97	\$46.78	20915270
225	\$54.65	\$54.47	17045060	260	\$46.73	\$46.13	35518250
226	\$54.51	\$54.42	15900080	261	\$47.43	\$46.85	86449630
227	\$54.69	\$53.68	46483390	262	\$50.04	\$48.70	49770590
228	\$54.45	\$54.30	26581850	263	\$48.80	\$48.41	18876580
229	\$54.57	\$54	25516130	264	\$48.98	\$48.96	25930810
230	\$53.96	\$53.34	20759890	265	\$49.38	\$49.11	29788730
231	\$53.65	\$52.88	18125150	266	\$48.84	\$48	17224030
232	\$52.56	\$51.97	16789600	267	\$48.89	\$48.60	18936240
233	\$52.99	\$52.77	17802960	268	\$49.13	\$48.71	22401790
234	\$53.65	\$53.63	14707390	269	\$48.83	\$48.46	19035810
235	\$53.99	\$53.85	20073560	270	\$49.29	\$48.10	21599330
236	\$54.10	\$53.75	21553130	271	\$48.65	\$47.75	27126590
237	\$53.68	\$53.44	18297780	272	\$48.57	\$47.96	22528800
238	\$53.18	\$52.92	18124460	273	\$48.03	\$47.80	22551790
239	\$53.38	\$52.76	17588260	274	\$48.00	\$47.10	22491020
240	\$53.37	\$53.01	16029950	275	\$47.57	\$45.84	35321630
241	\$53.61	\$53.06	20558900	276	\$46.28	\$46.15	31967690
242	\$52.98	\$52.70	31916620	277	\$47.47	\$45.96	18733890
243	\$51.84	\$51.32	16090480	278	\$47.48	\$47.09	20484740
244	\$51.73	\$51.63	20795430	279	\$47.74	\$46.51	24701860
245	\$51.71	\$51.35	15579990	280	\$46.38	\$45.62	28821130



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
281	\$46.20	\$44	28570180	316	\$48.25	\$48.13	30630070
282	\$45.17	\$44.63	19241560	317	\$48.47	\$46.95	28215450
283	\$46.65	\$45.75	52528650	318	\$48.56	\$48.29	33343570
284	\$46.31	\$45.41	37665490	319	\$47.88	\$45.53	43941200
285	\$47.73	\$47.49	39301200	320	\$46.75	\$46.50	35444440
286	\$48.45	\$47.47	28021790	321	\$46.44	\$45.59	67247680
287	\$48.41	\$47.86	26382210	322	\$44.94	\$43.54	47052220
288	\$48.76	\$47.89	26366290	323	\$44.36	\$44.16	33189380
289	\$48.58	\$48.26	19303560	324	\$45.09	\$44.18	38244220
290	\$48.64	\$48.58	31251120	325	\$45.34	\$44.50	34236640
291	\$48.59	\$47.90	20956500	326	\$45.18	\$45.02	28270420
292	\$47.46	\$46.28	22818690	327	\$45.75	\$45.65	26311200
293	\$48.06	\$47.94	33055310	328	\$46.33	\$46	23111280
294	\$48.39	\$46.80	31225600	329	\$46.01	\$44.90	21268670
295	\$49.60	\$49.52	33374960	330	\$45.14	\$44.63	21407880
296	\$50.50	\$50	29972310	331	\$45.34	\$45.27	31749480
297	\$49.32	\$47.95	43281210	332	\$45.61	\$45.15	43483080
298	\$48.56	\$48.49	23230250	333	\$46.12	\$46.06	41642550
299	\$48.88	\$47.94	27660080	334	\$47.62	\$47.18	23272260
300	\$48.22	\$46.94	27598840	335	\$47.35	\$46.96	18524970
301	\$47.46	\$47.27	19976140	336	\$48.34	\$48.23	28893760
302	\$47.07	\$46.69	8906879	337	\$48.86	\$48.84	28662300
303	\$47.88	\$47.85	18811570	338	\$49.72	\$48.62	38872430
304	\$48.12	\$47.40	29192730	339	\$48.88	\$47.51	40746390
305	\$49.15	\$48.61	26435640	340	\$47.22	\$46.79	23653550
306	\$49.05	\$47.77	31417980	341	\$47.75	\$45.62	46200440
307	\$48.24	\$47.01	28937860	342	\$46.17	\$45.91	15287150
308	\$47.97	\$47.85	25375230	343	\$46.30	\$46.04	23412530
309	\$48.25	\$46.99	29284040	344	\$46.39	\$46.22	22027790
310	\$47.80	\$47.71	24055970	345	\$47	\$46.31	17831910
311	\$49.03	\$48.79	21752760	346	\$47.42	\$47.12	52262680
312	\$49.29	\$48.49	26579920	347	\$47.36	\$46.78	21765110
313	\$48.75	\$47.63	29900000	348	\$46.44	\$46.09	15829560
314	\$47.92	\$47.80	23802430	349	\$46.53	\$45.58	20366080
315	\$48.25	\$47.34	23742310	350	\$45.90	\$45.29	17242530



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
351	\$45.97	\$45.78	22697430
352	\$45.85	\$45.68	18498420
353	\$44.98	\$44.56	29723230
354	\$46.03	\$46	32438230
355	\$46.76	\$46.56	21264970
356	\$47.48	\$46.91	25792810
357	\$47.96	\$47.76	19315070
358	\$48.12	\$47.75	20798820
359	\$48.43	\$48.38	15669670
360	\$48.56	\$47.96	16385680
361	\$48.83	\$48.52	16279490
362	\$48.86	\$48.70	13296010
363	\$48.79	\$48.55	13468210
364	\$48.39	\$48	17096430
365	\$47.89	\$47.29	17912840
366	\$47.17	\$46.86	31038380
367	\$47.35	\$47.25	23097140
368	\$47.85	\$46.66	18617550
369	\$47.12	\$47.05	24039470
370	\$47.39	\$47	26368390
371	\$47.95	\$47.75	24287200
372	\$48	\$47.78	26555070
373	\$48.76	\$48.67	22407870
374	\$49.24	\$48.92	17000380
375	\$49.17	\$48.51	28695820
376	\$50.60	\$49.86	23362970
377	\$50.13	\$49.60	21981650
378	\$49.83	\$49.43	28315470
379	\$49.31	\$48.77	26445230
380	\$49.78	\$49.50	22004190
381	\$49.52	\$48.50	26387650
382	\$48.91	\$48.06	28054040
383	\$48.68	\$48.24	36406210
384	\$47.99	\$47.63	35212830
385	\$49.20	\$48.76	67781420

OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
386	\$53.18	\$52.45	26848630
387	\$52.45	\$52.21	19756140
388	\$53.30	\$52.76	17786160
389	\$52.50	\$51.77	15867890
390	\$52.07	\$51.83	17380050
391	\$52.01	\$51.60	15688590
392	\$51.97	\$51.87	14641310
393	\$51.88	\$51.05	15702030
394	\$52.32	\$52.15	15893050
395	\$52.42	\$52.40	13747780
396	\$52.48	\$52.05	20288260
397	\$51.87	\$51.69	18642300
398	\$52.26	\$52.10	17323870
399	\$51.89	\$51.64	21343680
400	\$51.50	\$50.77	19959960
401	\$50.96	\$49.98	26106880
402	\$50.86	\$50.58	16327900
403	\$50.27	\$49.04	23741080
404	\$50.11	\$49.53	28902790
405	\$49.49	\$48.58	29052480
406	\$50.06	\$49.34	33280020
407	\$50.41	\$49.95	37512410
408	\$51.75	\$51.17	45320560
409	\$53.09	\$52.62	34664470
410	\$54.53	\$54.38	44370420



ANEXOS 2A. BASE DE DATOS SEGUNDO MODELO

OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
1	\$67.40	\$67.30	17408040	40	\$56.83	\$56.71	16157980
2	\$67.60	\$66.69	23401420	41	\$56.96	\$56.54	14467820
3	\$66.06	\$65.77	20970750	42	\$56.99	\$56.66	16477410
4	\$65.04	\$64.46	16654560	43	\$56.41	\$56.41	17675580
5	\$65.98	\$65.80	25268430	44	\$56.66	\$56.66	18882730
6	\$66.50	\$65.64	18522360	45	\$56.86	\$56.86	26613650
7	\$68.06	\$67.63	18076690	46	\$58.69	\$58.55	19166770
8	\$67.58	\$66.50	30915390	47	\$58.30	\$58.15	9481129
9	\$67.42	\$66.72	30764000	48	\$58.59	\$58.53	18185510
11	\$63.71	\$62.86	42778740	49	\$59.13	\$58.95	22381500
14	\$60	\$59.98	21803450	50	\$58.84	\$58	15012420
15	\$59.84	\$59.26	21365530	51	\$58.65	\$58.34	15690160
16	\$59.65	\$59.30	18498800	52	\$58.30	\$57.49	16139030
17	\$59.74	\$59.49	17092450	53	\$58.36	\$58.26	18559090
18	\$59.78	\$59.17	16455830	54	\$58.65	\$58.48	17062830
19	\$60.08	\$59.57	15201930	55	\$58.48	\$57.70	16101280
20	\$59.51	\$59.05	18818490	56	\$58.70	\$58.17	16112990
21	\$59.32	\$58.89	23138270	57	\$58.04	\$57.61	12141220
22	\$59.80	\$59.78	21960520	58	\$57.97	\$57.87	15409830
23	\$60.20	\$59.59	17759140	59	\$58.80	\$58.40	11954040
24	\$60.70	\$59.81	15298580	60	\$58.50	\$57.97	11236370
25	\$60.97	\$60.24	17713400	61	\$58.30	\$57.85	11890270
26	\$59.93	\$59.62	14435830	62	\$58.55	\$58	16962850
27	\$60.05	\$59.99	12751110	63	\$57.71	\$57.49	17721000
28	\$60.48	\$59.91	14086150	64	\$58.35	\$57.47	20865350
29	\$59.85	\$59.41	11486640	65	\$57.80	\$56.93	19067980
30	\$59.48	\$59.25	6313179	66	\$56.65	\$55.94	17953470
31	\$59.78	\$59.06	23593190	67	\$56.60	\$56.33	19249680
32	\$59.05	\$58.32	49970820	68	\$56.63	\$56.32	15248000
33	\$57.99	\$57.17	18869000	69	\$57.24	\$56.37	18155770
34	\$57.46	\$57.39	28273060	70	\$56.81	\$56.38	32307510
35	\$58.10	\$58	21656380	72	\$52.43	\$52.09	25654060
36	\$58.57	\$58.21	19135290	73	\$51.96	\$51.37	15961600
37	\$58.34	\$57.55	22795850	74	\$52.32	\$51.90	13347320
39	\$57.17	\$56.91	20051870	75	\$52.17	\$51.52	13120650



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
76	\$51.82	\$51.75	16975160	117	\$47.29	\$47.11	15915800
77	\$52.84	\$52.64	16918320	118	\$47.12	\$47.03	23115320
78	\$52.56	\$52.35	15173070	119	\$47.60	\$47.46	21404920
80	\$51.95	\$51.90	16425530	120	\$46.68	\$46.34	23335690
81	\$52.79	\$52	24167110	121	\$46.18	\$46.10	21992680
83	\$50.79	\$50.47	12189960	122	\$46.63	\$46.06	25650220
84	\$50.48	\$50.16	17884430	124	\$46.07	\$45.76	18494300
85	\$51.23	\$50.70	12722050	125	\$47	\$46.94	24975460
86	\$50.97	\$50.35	12748070	126	\$47.37	\$46.16	30643740
87	\$50.07	\$49.50	17235570	127	\$46.88	\$46.30	29440390
88	\$50.37	\$50.31	21437000	128	\$47.56	\$47.15	26119560
89	\$52.24	\$51.97	18487340	129	\$47.96	\$47.76	38936650
90	\$51.60	\$51.19	14059200	130	\$49.36	\$49.09	27881620
91	\$51.22	\$50.63	14846410	132	\$51.72	\$51.51	24169750
92	\$51.15	\$50.98	11482110	133	\$52.31	\$51.71	16487690
94	\$51.18	\$51.10	28472930	134	\$52.56	\$51.53	21897460
95	\$51.23	\$50.83	21574150	135	\$53.50	\$52.69	36401550
96	\$51.88	\$51.60	38632500	136	\$52.60	\$52.51	27765430
97	\$51.99	\$51.70	11849920	137	\$52.97	\$51.98	20896370
98	\$52.13	\$52.13	15029860	138	\$52	\$52	18413520
99	\$52.14	\$52.05	19888220	140	\$50.66	\$49.96	29141090
100	\$52.29	\$51.90	13354590	141	\$49.99	\$49.28	15679640
101	\$53	\$52.76	18010810	142	\$49.70	\$49.50	11928610
102	\$53.33	\$53	23308710	143	\$50	\$49.91	17131370
104	\$51.84	\$51.33	18532030	144	\$50.25	\$50	14345080
105	\$52.03	\$51.06	20749660	145	\$49.93	\$49	20148730
106	\$50.99	\$50.25	17067940	146	\$48.78	\$48.56	16615080
109	\$47.22	\$47.12	15010270	147	\$48.61	\$47.92	16032270
110	\$47.79	\$47.24	16923150	148	\$47.80	\$47.36	17853190
111	\$47.22	\$46.46	17806110	149	\$47.95	\$47.62	16564290
112	\$45.91	\$45.70	14895350	150	\$48.18	\$48.03	14339230
113	\$46.12	\$45.87	16927570	151	\$48.56	\$48.18	10739100
114	\$45.82	\$45.82	22086630	152	\$48.22	\$47.91	15384690
115	\$46.63	\$46.35	32817080	153	\$49.45	\$49.29	27505450
116	\$47.43	\$47.28	19785890	154	\$48.20	\$47.77	26646180



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
155	\$48.42	\$48.15	23062950	194	\$51.33	\$50.80	22655400
156	\$48.48	\$47.65	21886310	195	\$51.60	\$51.10	20467580
157	\$47.68	\$47.56	16665580	196	\$51.43	\$51	25617960
158	\$47.85	\$47.40	14479940	197	\$52.43	\$52.17	38546980
159	\$47.73	\$46.95	31096250	199	\$58.38	\$58.30	27494710
160	\$48.05	\$47.87	20126780	200	\$59.29	\$58.52	17603860
161	\$47.58	\$47.48	17436380	201	\$58.98	\$58.89	17016380
163	\$46.53	\$46.14	12643340	202	\$58.83	\$58.15	12313140
164	\$46.42	\$45.83	15381080	203	\$58.88	\$58.71	20014510
165	\$46.96	\$46.60	12267630	204	\$59.59	\$58.44	38572210
166	\$46.67	\$46.67	15900570	205	\$57.47	\$56.66	30057100
167	\$47.32	\$47.26	18415500	206	\$56.56	\$56.46	10986910
168	\$47.19	\$46.23	20806080	207	\$56.49	\$56.21	12639630
169	\$46.12	\$45.25	24370250	208	\$56.07	\$55.76	10807340
170	\$45.17	\$44.75	19076170	209	\$55.86	\$55.32	12841760
171	\$45.18	\$45.11	20802010	210	\$55.43	\$55.35	13204510
172	\$44.86	\$44.16	28094920	211	\$55.74	\$55.28	11751750
173	\$44.59	\$44.25	31155680	212	\$55.95	\$55.75	18919000
174	\$44.79	\$44.18	25532530	213	\$56.14	\$55.27	16421070
175	\$45.03	\$44.30	21618860	214	\$55.99	\$55.03	24439760
176	\$44.53	\$43.35	28450340	215	\$54.80	\$54.53	13507980
178	\$45.05	\$44.59	27168070	216	\$54.57	\$54.34	17216940
180	\$44.46	\$44.15	16932520	217	\$54.04	\$53.80	18568690
181	\$44.78	\$44.34	25543370	218	\$53.44	\$53.25	13479770
182	\$44.52	\$44	28674970	219	\$53.86	\$53.59	19185200
183	\$45.73	\$45.18	22500360	220	\$53.77	\$53.32	17709670
184	\$45.95	\$45.47	19954650	221	\$53.55	\$52.97	16638820
185	\$45.75	\$44.92	23407890	222	\$54.62	\$54.40	21329170
186	\$45.49	\$45.18	24706460	224	\$54.40	\$54.06	23330720
187	\$45.64	\$45.33	39091930	225	\$54.65	\$54.47	17045060
188	\$46.80	\$46.44	42522780	226	\$54.51	\$54.42	15900080
190	\$50.79	\$50.20	36812430	227	\$54.69	\$53.68	46483390
191	\$50.88	\$50.60	29713380	228	\$54.45	\$54.30	26581850
192	\$51.28	\$50.25	26245240	229	\$54.57	\$54	25516130
193	\$51.82	\$51.08	25656640	230	\$53.96	\$53.34	20759890



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
231	\$53.65	\$52.88	18125150	269	\$48.83	\$48.46	19035810
232	\$52.56	\$51.97	16789600	271	\$48.65	\$47.75	27126590
233	\$52.99	\$52.77	17802960	272	\$48.57	\$47.96	22528800
234	\$53.65	\$53.63	14707390	273	\$48.03	\$47.80	22551790
235	\$53.99	\$53.85	20073560	274	\$48.00	\$47.10	22491020
236	\$54.10	\$53.75	21553130	276	\$46.28	\$46.15	31967690
237	\$53.68	\$53.44	18297780	278	\$47.48	\$47.09	20484740
238	\$53.18	\$52.92	18124460	280	\$46.38	\$45.62	28821130
239	\$53.38	\$52.76	17588260	282	\$45.17	\$44.63	19241560
240	\$53.37	\$53.01	16029950	283	\$46.65	\$45.75	52528650
241	\$53.61	\$53.06	20558900	284	\$46.31	\$45.41	37665490
242	\$52.98	\$52.70	31916620	285	\$47.73	\$47.49	39301200
243	\$51.84	\$51.32	16090480	286	\$48.45	\$47.47	28021790
244	\$51.73	\$51.63	20795430	287	\$48.41	\$47.86	26382210
245	\$51.71	\$51.35	15579990	288	\$48.76	\$47.89	26366290
246	\$51.67	\$51.25	22512380	289	\$48.58	\$48.26	19303560
247	\$51.04	\$50.41	17530450	290	\$48.64	\$48.58	31251120
248	\$50.81	\$50.13	17501310	291	\$48.59	\$47.90	20956500
249	\$50.14	\$49.36	18039220	293	\$48.06	\$47.94	33055310
250	\$49.32	\$49.02	17998810	295	\$49.60	\$49.52	33374960
251	\$49.02	\$48.77	18848690	296	\$50.50	\$50	29972310
252	\$49.85	\$49.50	27807690	297	\$49.32	\$47.95	43281210
253	\$50.64	\$49.95	21642300	298	\$48.56	\$48.49	23230250
255	\$49.24	\$48.62	20851710	299	\$48.88	\$47.94	27660080
257	\$47.18	\$46.70	41457610	301	\$47.46	\$47.27	19976140
258	\$47.68	\$46.68	26746480	302	\$47.07	\$46.69	8906879
259	\$46.97	\$46.78	20915270	303	\$47.88	\$47.85	18811570
260	\$46.73	\$46.13	35518250	304	\$48.12	\$47.40	29192730
262	\$50.04	\$48.70	49770590	305	\$49.15	\$48.61	26435640
263	\$48.80	\$48.41	18876580	308	\$47.97	\$47.85	25375230
264	\$48.98	\$48.96	25930810	310	\$47.80	\$47.71	24055970
265	\$49.38	\$49.11	29788730	311	\$49.03	\$48.79	21752760
266	\$48.84	\$48	17224030	312	\$49.29	\$48.49	26579920
267	\$48.89	\$48.60	18936240	313	\$48.75	\$47.63	29900000
268	\$49.13	\$48.71	22401790	314	\$47.92	\$47.80	23802430



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
315	\$48.25	\$47.34	23742310	357	\$47.96	\$47.76	19315070
316	\$48.25	\$48.13	30630070	358	\$48.12	\$47.75	20798820
318	\$48.56	\$48.29	33343570	359	\$48.43	\$48.38	15669670
320	\$46.75	\$46.50	35444440	360	\$48.56	\$47.96	16385680
321	\$46.44	\$45.59	67247680	361	\$48.83	\$48.52	16279490
322	\$44.94	\$43.54	47052220	362	\$48.86	\$48.70	13296010
323	\$44.36	\$44.16	33189380	363	\$48.79	\$48.55	13468210
324	\$45.09	\$44.18	38244220	364	\$48.39	\$48	17096430
325	\$45.34	\$44.50	34236640	365	\$47.89	\$47.29	17912840
326	\$45.18	\$45.02	28270420	366	\$47.17	\$46.86	31038380
327	\$45.75	\$45.65	26311200	367	\$47.35	\$47.25	23097140
328	\$46.33	\$46	23111280	369	\$47.12	\$47.05	24039470
330	\$45.14	\$44.63	21407880	370	\$47.39	\$47	26368390
331	\$45.34	\$45.27	31749480	371	\$47.95	\$47.75	24287200
332	\$45.61	\$45.15	43483080	372	\$48	\$47.78	26555070
334	\$47.62	\$47.18	23272260	373	\$48.76	\$48.67	22407870
335	\$47.35	\$46.96	18524970	374	\$49.24	\$48.92	17000380
336	\$48.34	\$48.23	28893760	375	\$49.17	\$48.51	28695820
337	\$48.86	\$48.84	28662300	376	\$50.60	\$49.86	23362970
338	\$49.72	\$48.62	38872430	377	\$50.13	\$49.60	21981650
339	\$48.88	\$47.51	40746390	378	\$49.83	\$49.43	28315470
340	\$47.22	\$46.79	23653550	379	\$49.31	\$48.77	26445230
342	\$46.17	\$45.91	15287150	380	\$49.78	\$49.50	22004190
343	\$46.30	\$46.04	23412530	381	\$49.52	\$48.50	26387650
344	\$46.39	\$46.22	22027790	382	\$48.91	\$48.06	28054040
345	\$47	\$46.31	17831910	383	\$48.68	\$48.24	36406210
347	\$47.36	\$46.78	21765110	384	\$47.99	\$47.63	35212830
348	\$46.44	\$46.09	15829560	386	\$53.18	\$52.45	26848630
349	\$46.53	\$45.58	20366080	387	\$52.45	\$52.21	19756140
350	\$45.90	\$45.29	17242530	388	\$53.30	\$52.76	17786160
351	\$45.97	\$45.78	22697430	389	\$52.50	\$51.77	15867890
352	\$45.85	\$45.68	18498420	390	\$52.07	\$51.83	17380050
353	\$44.98	\$44.56	29723230	391	\$52.01	\$51.60	15688590
355	\$46.76	\$46.56	21264970	392	\$51.97	\$51.87	14641310
356	\$47.48	\$46.91	25792810	393	\$51.88	\$51.05	15702030



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
394	\$52.32	\$52.15	15893050
395	\$52.42	\$52.40	13747780
396	\$52.48	\$52.05	20288260
397	\$51.87	\$51.69	18642300
398	\$52.26	\$52.10	17323870
399	\$51.89	\$51.64	21343680
400	\$51.50	\$50.77	19959960
401	\$50.96	\$49.98	26106880
402	\$50.86	\$50.58	16327900
404	\$50.11	\$49.53	28902790
405	\$49.49	\$48.58	29052480
406	\$50.06	\$49.34	33280020
407	\$50.41	\$49.95	37512410
408	\$51.75	\$51.17	45320560
409	\$53.09	\$52.62	34664470



ANEXOS 2B. BASE DE DATOS TERCER MODELO

OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
1	\$67.40	\$67.30	17408040	49	\$59.13	\$58.95	22381500
2	\$67.60	\$66.69	23401420	50	\$58.84	\$58	15012420
3	\$66.06	\$65.77	20970750	51	\$58.65	\$58.34	15690160
4	\$65.04	\$64.46	16654560	52	\$58.30	\$57.49	16139030
5	\$65.98	\$65.80	25268430	53	\$58.36	\$58.26	18559090
7	\$68.06	\$67.63	18076690	54	\$58.65	\$58.48	17062830
9	\$67.42	\$66.72	30764000	55	\$58.48	\$57.70	16101280
14	\$60	\$59.98	21803450	56	\$58.70	\$58.17	16112990
15	\$59.84	\$59.26	21365530	57	\$58.04	\$57.61	12141220
16	\$59.65	\$59.30	18498800	58	\$57.97	\$57.87	15409830
17	\$59.74	\$59.49	17092450	59	\$58.80	\$58.40	11954040
18	\$59.78	\$59.17	16455830	62	\$58.55	\$58	16962850
20	\$59.51	\$59.05	18818490	63	\$57.71	\$57.49	17721000
22	\$59.80	\$59.78	21960520	64	\$58.35	\$57.47	20865350
23	\$60.20	\$59.59	17759140	66	\$56.65	\$55.94	17953470
24	\$60.70	\$59.81	15298580	67	\$56.60	\$56.33	19249680
25	\$60.97	\$60.24	17713400	68	\$56.63	\$56.32	15248000
26	\$59.93	\$59.62	14435830	69	\$57.24	\$56.37	18155770
27	\$60.05	\$59.99	12751110	70	\$56.81	\$56.38	32307510
28	\$60.48	\$59.91	14086150	72	\$52.43	\$52.09	25654060
29	\$59.85	\$59.41	11486640	73	\$51.96	\$51.37	15961600
30	\$59.48	\$59.25	6313179	74	\$52.32	\$51.90	13347320
32	\$59.05	\$58.32	49970820	75	\$52.17	\$51.52	13120650
33	\$57.99	\$57.17	18869000	76	\$51.82	\$51.75	16975160
34	\$57.46	\$57.39	28273060	77	\$52.84	\$52.64	16918320
35	\$58.10	\$58	21656380	78	\$52.56	\$52.35	15173070
36	\$58.57	\$58.21	19135290	80	\$51.95	\$51.90	16425530
37	\$58.34	\$57.55	22795850	81	\$52.79	\$52	24167110
39	\$57.17	\$56.91	20051870	83	\$50.79	\$50.47	12189960
40	\$56.83	\$56.71	16157980	84	\$50.48	\$50.16	17884430
42	\$56.99	\$56.66	16477410	85	\$51.23	\$50.70	12722050
43	\$56.41	\$56.41	17675580	87	\$50.07	\$49.50	17235570
44	\$56.66	\$56.66	18882730	88	\$50.37	\$50.31	21437000
45	\$56.86	\$56.86	26613650	89	\$52.24	\$51.97	18487340
47	\$58.30	\$58.15	9481129	90	\$51.60	\$51.19	14059200



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
91	\$51.22	\$50.63	14846410	137	\$52.97	\$51.98	20896370
92	\$51.15	\$50.98	11482110	138	\$52	\$52	18413520
94	\$51.18	\$51.10	28472930	140	\$50.66	\$49.96	29141090
95	\$51.23	\$50.83	21574150	141	\$49.99	\$49.28	15679640
97	\$51.99	\$51.70	11849920	142	\$49.70	\$49.50	11928610
98	\$52.13	\$52.13	15029860	143	\$50	\$49.91	17131370
99	\$52.14	\$52.05	19888220	144	\$50.25	\$50	14345080
100	\$52.29	\$51.90	13354590	145	\$49.93	\$49	20148730
101	\$53	\$52.76	18010810	146	\$48.78	\$48.56	16615080
102	\$53.33	\$53	23308710	147	\$48.61	\$47.92	16032270
104	\$51.84	\$51.33	18532030	148	\$47.80	\$47.36	17853190
105	\$52.03	\$51.06	20749660	149	\$47.95	\$47.62	16564290
106	\$50.99	\$50.25	17067940	150	\$48.18	\$48.03	14339230
109	\$47.22	\$47.12	15010270	151	\$48.56	\$48.18	10739100
110	\$47.79	\$47.24	16923150	152	\$48.22	\$47.91	15384690
111	\$47.22	\$46.46	17806110	154	\$48.20	\$47.77	26646180
112	\$45.91	\$45.70	14895350	155	\$48.42	\$48.15	23062950
113	\$46.12	\$45.87	16927570	156	\$48.48	\$47.65	21886310
115	\$46.63	\$46.35	32817080	157	\$47.68	\$47.56	16665580
116	\$47.43	\$47.28	19785890	158	\$47.85	\$47.40	14479940
118	\$47.12	\$47.03	23115320	159	\$47.73	\$46.95	31096250
119	\$47.60	\$47.46	21404920	160	\$48.05	\$47.87	20126780
120	\$46.68	\$46.34	23335690	163	\$46.53	\$46.14	12643340
122	\$46.63	\$46.06	25650220	164	\$46.42	\$45.83	15381080
125	\$47	\$46.94	24975460	165	\$46.96	\$46.60	12267630
126	\$47.37	\$46.16	30643740	166	\$46.67	\$46.67	15900570
127	\$46.88	\$46.30	29440390	167	\$47.32	\$47.26	18415500
128	\$47.56	\$47.15	26119560	168	\$47.19	\$46.23	20806080
129	\$47.96	\$47.76	38936650	169	\$46.12	\$45.25	24370250
130	\$49.36	\$49.09	27881620	170	\$45.17	\$44.75	19076170
132	\$51.72	\$51.51	24169750	171	\$45.18	\$45.11	20802010
133	\$52.31	\$51.71	16487690	172	\$44.86	\$44.16	28094920
134	\$52.56	\$51.53	21897460	173	\$44.59	\$44.25	31155680
135	\$53.50	\$52.69	36401550	175	\$45.03	\$44.30	21618860
136	\$52.60	\$52.51	27765430	176	\$44.53	\$43.35	28450340



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
178	\$45.05	\$44.59	27168070	220	\$53.77	\$53.32	17709670
181	\$44.78	\$44.34	25543370	221	\$53.55	\$52.97	16638820
182	\$44.52	\$44	28674970	222	\$54.62	\$54.40	21329170
183	\$45.73	\$45.18	22500360	224	\$54.40	\$54.06	23330720
184	\$45.95	\$45.47	19954650	225	\$54.65	\$54.47	17045060
186	\$45.49	\$45.18	24706460	226	\$54.51	\$54.42	15900080
187	\$45.64	\$45.33	39091930	227	\$54.69	\$53.68	46483390
188	\$46.80	\$46.44	42522780	228	\$54.45	\$54.30	26581850
190	\$50.79	\$50.20	36812430	229	\$54.57	\$54	25516130
191	\$50.88	\$50.60	29713380	230	\$53.96	\$53.34	20759890
192	\$51.28	\$50.25	26245240	231	\$53.65	\$52.88	18125150
193	\$51.82	\$51.08	25656640	232	\$52.56	\$51.97	16789600
196	\$51.43	\$51	25617960	233	\$52.99	\$52.77	17802960
197	\$52.43	\$52.17	38546980	234	\$53.65	\$53.63	14707390
199	\$58.38	\$58.30	27494710	235	\$53.99	\$53.85	20073560
200	\$59.29	\$58.52	17603860	237	\$53.68	\$53.44	18297780
201	\$58.98	\$58.89	17016380	238	\$53.18	\$52.92	18124460
202	\$58.83	\$58.15	12313140	240	\$53.37	\$53.01	16029950
203	\$58.88	\$58.71	20014510	242	\$52.98	\$52.70	31916620
204	\$59.59	\$58.44	38572210	244	\$51.73	\$51.63	20795430
205	\$57.47	\$56.66	30057100	245	\$51.71	\$51.35	15579990
206	\$56.56	\$56.46	10986910	246	\$51.67	\$51.25	22512380
207	\$56.49	\$56.21	12639630	247	\$51.04	\$50.41	17530450
208	\$56.07	\$55.76	10807340	248	\$50.81	\$50.13	17501310
209	\$55.86	\$55.32	12841760	249	\$50.14	\$49.36	18039220
210	\$55.43	\$55.35	13204510	251	\$49.02	\$48.77	18848690
211	\$55.74	\$55.28	11751750	252	\$49.85	\$49.50	27807690
212	\$55.95	\$55.75	18919000	253	\$50.64	\$49.95	21642300
213	\$56.14	\$55.27	16421070	255	\$49.24	\$48.62	20851710
214	\$55.99	\$55.03	24439760	257	\$47.18	\$46.70	41457610
215	\$54.80	\$54.53	13507980	259	\$46.97	\$46.78	20915270
216	\$54.57	\$54.34	17216940	260	\$46.73	\$46.13	35518250
217	\$54.04	\$53.80	18568690	263	\$48.80	\$48.41	18876580
218	\$53.44	\$53.25	13479770	266	\$48.84	\$48	17224030
219	\$53.86	\$53.59	19185200	268	\$49.13	\$48.71	22401790



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN	OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
271	\$48.65	\$47.75	27126590	321	\$46.44	\$45.59	67247680
272	\$48.57	\$47.96	22528800	322	\$44.94	\$43.54	47052220
273	\$48.03	\$47.80	22551790	323	\$44.36	\$44.16	33189380
274	\$48.00	\$47.10	22491020	324	\$45.09	\$44.18	38244220
276	\$46.28	\$46.15	31967690	325	\$45.34	\$44.50	34236640
278	\$47.48	\$47.09	20484740	326	\$45.18	\$45.02	28270420
280	\$46.38	\$45.62	28821130	327	\$45.75	\$45.65	26311200
282	\$45.17	\$44.63	19241560	328	\$46.33	\$46	23111280
283	\$46.65	\$45.75	52528650	330	\$45.14	\$44.63	21407880
284	\$46.31	\$45.41	37665490	331	\$45.34	\$45.27	31749480
285	\$47.73	\$47.49	39301200	332	\$45.61	\$45.15	43483080
287	\$48.41	\$47.86	26382210	334	\$47.62	\$47.18	23272260
288	\$48.76	\$47.89	26366290	335	\$47.35	\$46.96	18524970
289	\$48.58	\$48.26	19303560	336	\$48.34	\$48.23	28893760
290	\$48.64	\$48.58	31251120	337	\$48.86	\$48.84	28662300
291	\$48.59	\$47.90	20956500	338	\$49.72	\$48.62	38872430
293	\$48.06	\$47.94	33055310	340	\$47.22	\$46.79	23653550
295	\$49.60	\$49.52	33374960	342	\$46.17	\$45.91	15287150
296	\$50.50	\$50	29972310	343	\$46.30	\$46.04	23412530
297	\$49.32	\$47.95	43281210	344	\$46.39	\$46.22	22027790
298	\$48.56	\$48.49	23230250	345	\$47	\$46.31	17831910
302	\$47.07	\$46.69	8906879	347	\$47.36	\$46.78	21765110
303	\$47.88	\$47.85	18811570	348	\$46.44	\$46.09	15829560
304	\$48.12	\$47.40	29192730	349	\$46.53	\$45.58	20366080
305	\$49.15	\$48.61	26435640	351	\$45.97	\$45.78	22697430
308	\$47.97	\$47.85	25375230	352	\$45.85	\$45.68	18498420
310	\$47.80	\$47.71	24055970	353	\$44.98	\$44.56	29723230
311	\$49.03	\$48.79	21752760	355	\$46.76	\$46.56	21264970
312	\$49.29	\$48.49	26579920	356	\$47.48	\$46.91	25792810
313	\$48.75	\$47.63	29900000	357	\$47.96	\$47.76	19315070
314	\$47.92	\$47.80	23802430	359	\$48.43	\$48.38	15669670
315	\$48.25	\$47.34	23742310	360	\$48.56	\$47.96	16385680
316	\$48.25	\$48.13	30630070	361	\$48.83	\$48.52	16279490
318	\$48.56	\$48.29	33343570	362	\$48.86	\$48.70	13296010
320	\$46.75	\$46.50	35444440	363	\$48.79	\$48.55	13468210



OBSERVACIÓN	PRECIO MAXIMO	PRECIO APERTURA	VOLUMEN
364	\$48.39	\$48	17096430
365	\$47.89	\$47.29	17912840
366	\$47.17	\$46.86	31038380
367	\$47.35	\$47.25	23097140
369	\$47.12	\$47.05	24039470
370	\$47.39	\$47	26368390
371	\$47.95	\$47.75	24287200
372	\$48	\$47.78	26555070
373	\$48.76	\$48.67	22407870
374	\$49.24	\$48.92	17000380
375	\$49.17	\$48.51	28695820
376	\$50.60	\$49.86	23362970
377	\$50.13	\$49.60	21981650
378	\$49.83	\$49.43	28315470
379	\$49.31	\$48.77	26445230
380	\$49.78	\$49.50	22004190
381	\$49.52	\$48.50	26387650
382	\$48.91	\$48.06	28054040
383	\$48.68	\$48.24	36406210
384	\$47.99	\$47.63	35212830
386	\$53.18	\$52.45	26848630
387	\$52.45	\$52.21	19756140
388	\$53.30	\$52.76	17786160
389	\$52.50	\$51.77	15867890
390	\$52.07	\$51.83	17380050
391	\$52.01	\$51.60	15688590
392	\$51.97	\$51.87	14641310
393	\$51.88	\$51.05	15702030
394	\$52.32	\$52.15	15893050
395	\$52.42	\$52.40	13747780
396	\$52.48	\$52.05	20288260
397	\$51.87	\$51.69	18642300
398	\$52.26	\$52.10	17323870
399	\$51.89	\$51.64	21343680
400	\$51.50	\$50.77	19959960
401	\$50.96	\$49.98	26106880
402	\$50.86	\$50.58	16327900
404	\$50.11	\$49.53	28902790
405	\$49.49	\$48.58	29052480
406	\$50.06	\$49.34	33280020
407	\$50.41	\$49.95	37512410
408	\$51.75	\$51.17	45320560
409	\$53.09	\$52.62	34664470



ANEXOS 3. BASE DE DATOS OBTENIDO DE NASDAQ

<https://www.nasdaq.com/es/market-activity/stocks/intc>



Nasdaq ACTIVIDAD DEL MERCADO SOLUCIONES

Datos clave

Cambio	NASDAQ-GS	Cap. de mercado	248,364,689,200
Sector	Technology	Coefficiente P/E	12.37
Industria	Semiconductors	Índice P/E estimado a 1 año	12.35
Objetivo a 1 año	\$60.00	Beneficios por acción (BPA)	\$4.94
Alto/bajo de hoy	\$61.40/\$60.38	Dividendo anualizado	\$1.39
Volumen de acciones	3,318,963	Fecha ex dividendo	Feb 4, 2021
Volumen Promedio	39,177,993	Fecha de pago de dividendo	Mar 1, 2021
Cierre anterior	\$61.12	Rendimiento actual	2.29%
52 semanas alto/bajo	\$65.11/\$43.61	Beta	0.68