

UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN |
FACULTAD DE CIENCIAS FÍSICO MATEMÁTICAS
MAESTRÍA EN CIENCIA DE DATOS | METODOS ESTADÍSTICOS BÁSICOS |
MET. ALEJANDRA GUADALUPE CERDA RUIZ

ACT. ERICK ADRIÁN GARZA TAMEZ | 1812629

San Nicolas de los Garza a 4 de Diciembre de 2022

Introducción

A lo largo de este documento se mostrarán estudios estadísticos para conocer como influyen diversos indicadores económicos en el precio de cotización de la acción de Apple. Las variables que nos interesan saber si influyen sobre esta son por ejemplo la inflación en Estados Unidos, su volumen operado, el precio del dólar en pesos mexicanos y la cotización de los índices S&P 500 y el Dow Jones, los cuales son los principales indicadores económicos que repercuten en la sociedad.

Para fines prácticos de este proyecto los datos a utilizar serán tomados del 04 de octubre de 2021 hasta el 30 de septiembre de 2022, teniendo un total de 251 registros para su posterior estudio. Estos datos fueron obtenidos en el sitio oficial de Yahoo Finance el cual se anexa cada uno de los links en la bibliografía del reporte.

Descripción de las Variables

Para este reporte se utilizará la variable de respuesta (variable de interés) la cual sigue el **precio de la cotización de la acción de Apple** con una denominación en dólares. Esta es una variable continua y sigue una escala de intervalo.

Dicho esto, nos interesa saber como influyen nuestras variables predictivas (variables independientes) en nuestra variable de respuesta. Para esto nos concentramos en 4 indicadores económicos importantes de Estados Unidos y el volumen diario operado de la acción, esto para afirmar si estas variables realmente influyen en el precio de la acción de Apple.

Nuestras variables predictivas son las siguientes:

- **Volumen Operado en Millones de Dólares:** Esta variable representa el volumen diario que se opera en el transcurso del día de la acción de Apple. Esta es una variable continua y sigue una escala de intervalo.
- **S&P 500:** Este es el índice Standard & Poor's 500, también conocido como S&P 500, es uno de los índices bursátiles más importantes de Estados Unidos. Al S&P 500 se lo considera el índice más representativo de la situación real del mercado. Esto se debe principalmente a que contiene a las 500 empresas más importantes de Estados Unidos, por lo que su movimiento da a conocer de manera general la situación actual de la economía estadounidense y global. Esta es una variable continua y sigue una escala de intervalo.
- **Inflación en Estados Unidos:** Esta variable indica el nivel de inflación en Estados Unidos, la inflación es el aumento sostenido y generalizado de los precios, de los bienes y servicios de una economía a lo largo del tiempo. Esta variable nos podría indicar que tanto poder adquisitivo se contaba en cada uno de los días de la cotización de Apple. Por teoría se podría decir que a mayor inflación esperas que haya disminución en el volumen operado e incluso un decremento en el precio de la acción. Esta es una variable continua y sigue una escala de intervalo.
- **Dow Jones Industrial (DJI):** El Dow Jones Industrial Average mide el desempeño de las 30 mayores empresas que cotizan en el mercado bursátil de Estados Unidos. Al igual que el S&P 500 este es un índice que refleja la salud financiera del mercado estadounidense. Esta variable es continua y sigue una escala de intervalo.

- **Precio del Dólar:** Esta variable es considerada para medir la variación entre el precio del dólar en pesos mexicanos y ver si influye directamente en el precio de la acción de Apple. ¿Podríamos afirmar que si aumenta el precio del dólar también aumenta el precio de la acción? Este tipo de preguntas son las que nos interesan conocer y desarrollar durante el proyecto. Esta es una variable continua y tiene una escala de intervalo.

A continuación se mostrará la cabecera y la cola de los datos a trabajar:

Fecha	Apple	volumen_en_millones	SP500	inflacion_usa	DJI	precio_dolar
2021-10-04	139.14	98.3220	428.64	6.2	28725.51	20.41820
2021-10-05	141.11	80.8611	433.10	6.2	29225.61	20.50160
2021-10-06	142.00	83.2211	434.90	6.2	29683.74	20.57020
2021-10-07	143.29	61.7327	438.66	6.2	29134.99	20.55750
2021-10-08	142.90	58.7732	437.86	6.2	29260.81	20.65413
2021-10-11	142.81	64.4522	434.69	6.2	29590.41	20.72315

Fecha	Apple	volumen_en_millones	SP500	inflacion_usa	DJI	precio_dolar
2022-09-23	150.43	95.9392	367.95	8.3	34496.06	19.85910
2022-09-26	150.77	93.3394	364.31	8.3	34746.25	19.83002
2022-09-27	151.76	84.4427	363.38	8.3	34754.94	20.03190
2022-09-28	149.84	146.6914	370.53	8.3	34416.99	19.95160
2022-09-29	142.48	128.1382	362.79	8.3	34314.67	20.08100
2022-09-30	138.20	124.7054	357.18	8.3	34002.92	20.03022

Alcances y Problemáticas a Resolver

El planteamiento va muy orientado a conocer el verdadero impacto que tienen las variables antes descritas sobre el precio de la acción de Apple. Algunas de las preguntas que nos gustaría resolver son las siguientes:

- Si aumenta el S&P 500, ¿cómo se ve afecta la acción de Apple?
- En un escenario de alta inflación, ¿la acción de Apple sufre decrementos?
- Con una cotización “cara” del precio del dólar, ¿aumenta el precio de la acción de Apple?

En sí, nos queremos enfocar a la resolución de escenarios atípicos como los que hemos vivido en los últimos años, logrando así un entendimiento más profundo sobre el desarrollo de la compañía en este tipo de circunstancias.

Análisis Estadístico

Para dar resolución a lo antes mencionado se trabajarán los datos y se aplicará un análisis estadístico en ellos los cuales incluyen un análisis exploratorio de los datos, regresión lineal múltiple, pruebas de bondad de ajuste y pruebas de hipótesis. En cada uno de estos procesos se espera ir conociendo con mayor profundidad las variables para poder llegar a conclusiones concretas sobre las problemáticas antes propuestas. Para realizar este análisis nos apoyaremos de la herramienta “**R Studio**”.

Análisis Exploratorio de los Datos

Para este apartado se realizará el análisis descriptivo tanto para las variables predictivas como la variable de respuesta, incluyendo tablas de frecuencias y gráficas tales como histogramas y boxplot.

Variable de Respuesta

Para conocer un poco mejor nuestra variable de respuesta es necesario conocer su tabla de frecuencias la cual nos indica su valor mínimo y máximo, como también la dispersión de sus datos respecto a la media y asimilar a que tipo de distribución se parecen sus datos.

A continuación se muestra la tabla de frecuencias del Precio de la Acción de Apple:

Lower	Upper	Main	Frequency	Percentage	CF	CPF
130	135	132.5	4	1.6	4	1.6
135	140	137.5	13	5.2	17	6.8
140	145	142.5	22	8.8	39	15.5
145	150	147.5	32	12.7	71	28.3
150	155	152.5	31	12.4	102	40.6
155	160	157.5	28	11.2	130	51.8
160	165	162.5	31	12.4	161	64.1
165	170	167.5	30	12.0	191	76.1
170	175	172.5	38	15.1	229	91.2
175	180	177.5	20	8.0	249	99.2
180	185	182.5	2	0.8	251	100.0

De lo anterior podemos observar que nuestro nivel más bajo es 130 y el máximo es de 185 dólares por acción. Para ver estos datos de una manera más sencilla realizaremos un histograma y un gráfico de Boxplot.

-
-
-
-
-

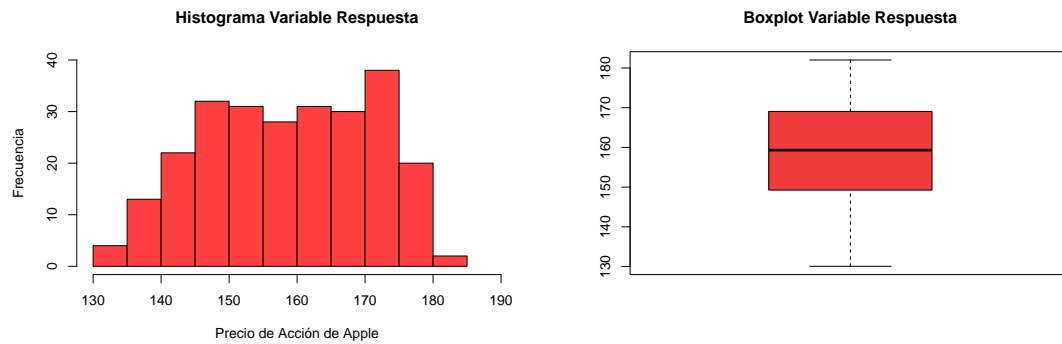


Figura 1: Histograma y Boxplot Precio Acción de Apple

Con la información mostrada en las figuras anteriores podemos identificar a manera sospechosa que nuestros datos tienden a comportarse de una manera **Lognormal**. Esto mismo por la propia naturaleza de los datos, ya que consisten en el precio de una acción del mercado bursátil. También podemos observar que la mayoría de los datos se concentran en los niveles entre 145 y 170 dólares por acción.

A continuación anexamos la distribución de datos a manera lognormal:

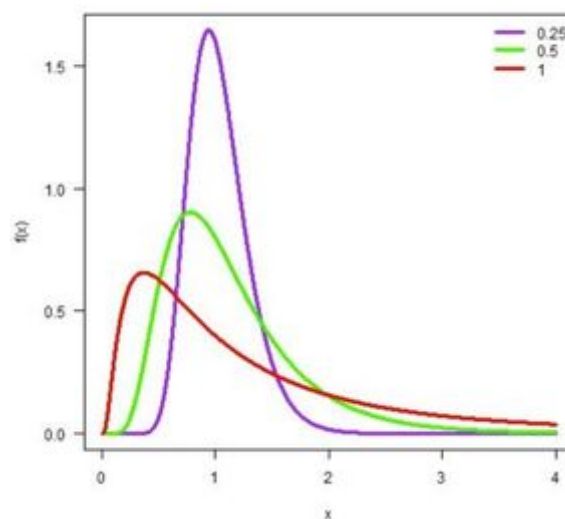


Figura 2: Distribución Lognormal

Variable Predictiva Volumen Operado

A continuación se muestra la tabla de frecuencias del Volumen Operado en Millones de Dólares:

Lower	Upper	Main	Frequency	Percentage	CF	CPF
40	60	50	24	9.6	24	9.6
60	80	70	86	34.3	110	43.8
80	100	90	75	29.9	185	73.7
100	120	110	32	12.7	217	86.5
120	140	130	22	8.8	239	95.2
140	160	150	6	2.4	245	97.6
160	180	170	4	1.6	249	99.2

Lower	Upper	Main	Frequency	Percentage	CF	CPF
180	200	190	2	0.8	251	100.0

De lo anterior podemos observar que el volumen más bajo operado en un día es de 40 millones de dólares y el máximo es de 200 millones de dólares. Para ver estos datos de una manera más sencilla realizaremos un histograma y un gráfico de Boxplot.

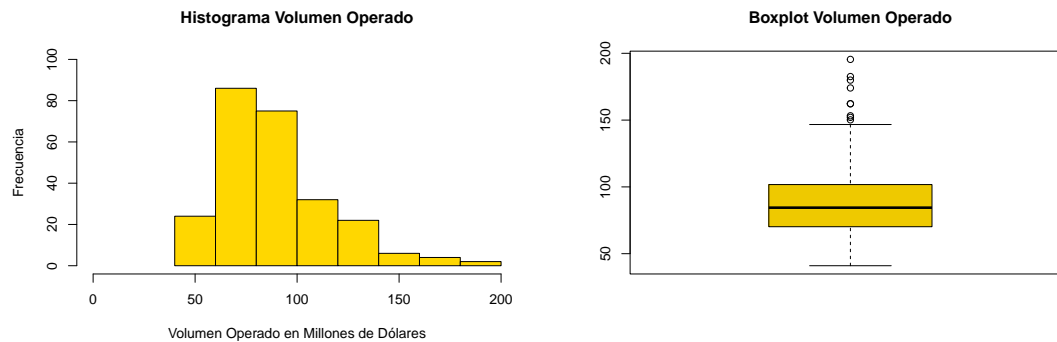


Figura 3: Histograma y Boxplot Volumen Operado

Con la información mostrada en las figuras anteriores podemos identificar a manera sospechosa que nuestros datos tienden a comportarse de una manera **Lognormal** la cual podemos ver su distribución en la figura 2. Aunado a esto, podemos detectar datos atípicos que tienden a acercarse hasta los 200 millones de dólares. También podemos observar que la mayoría de los datos se concentran en los niveles entre 60 y 100 millones de dólares operados por día.

Variable Predictiva S&P 500

A continuación se muestra la tabla de frecuencias del S&P 500:

Lower	Upper	Main	Frequency	Percentage	CF	CPF
350	360	355	1	0.4	1	0.4
360	370	365	6	2.4	7	2.8
370	380	375	12	4.8	19	7.6
380	390	385	18	7.2	37	14.7
390	400	395	24	9.6	61	24.3
400	410	405	13	5.2	74	29.5
410	420	415	31	12.4	105	41.8
420	430	425	19	7.6	124	49.4
430	440	435	26	10.4	150	59.8
440	450	445	24	9.6	174	69.3
450	460	455	31	12.4	205	81.7
460	470	465	35	13.9	240	95.6
470	480	475	11	4.4	251	100.0

De lo anterior podemos observar que la cotización mas baja se encuentra en los niveles de 350 puntos y el máximo a niveles de 480 puntos. Para ver estos datos de una manera más sencilla realizaremos un histograma y un gráfico de Boxplot.

-
-
-
-

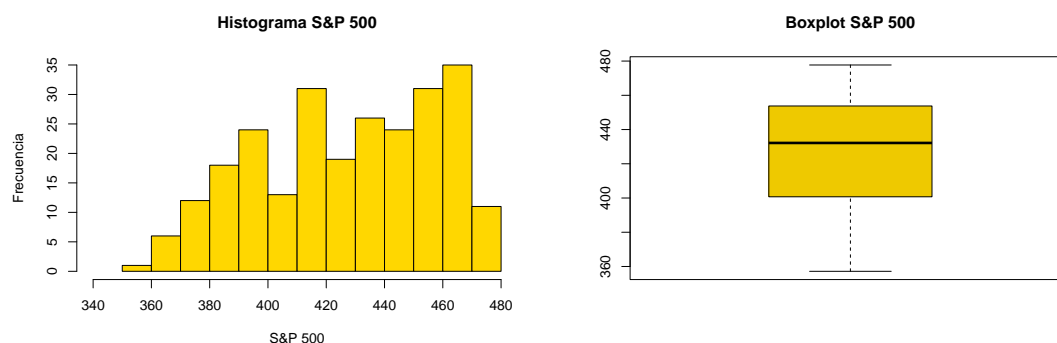


Figura 4: Histograma y Boxplot SP 500

Con la información mostrada en las figuras anteriores podemos identificar a manera sospechosa que nuestros datos tienden a comportarse de una manera **Lognormal** la cual podemos ver su distribución en la figura 2. A su vez, podemos detectar que los datos parecen demostrar una inclinación hacia la derecha del histograma, lo que nos puede indicar que se concentra la mayor parte de datos a niveles superiores de 400 puntos del S&P 500, lo cual lo podemos confirmar con nuestro Boxplot.

Variable Predictiva Inflación en Estados Unidos

A continuación se muestra la tabla de frecuencias de la inflación en Estados Unidos:

Lower	Upper	Main	Frequency	Percentage	CF	CPF
6.0	6.5	6.25	20	8.0	20	8.0
6.5	7.0	6.75	43	17.1	63	25.1
7.0	7.5	7.25	20	8.0	83	33.1
7.5	8.0	7.75	18	7.2	101	40.2
8.0	8.5	8.25	108	43.0	209	83.3
8.5	9.0	8.75	21	8.4	230	91.6
9.0	9.5	9.25	21	8.4	251	100.0

De lo anterior podemos observar que el nivel mas bajo de inflación se encuentra en niveles de 6% y el máximo en niveles de 9.5% anual. Para ver estos datos de una manera más sencilla realizaremos un histograma y un gráfico de Boxplot.

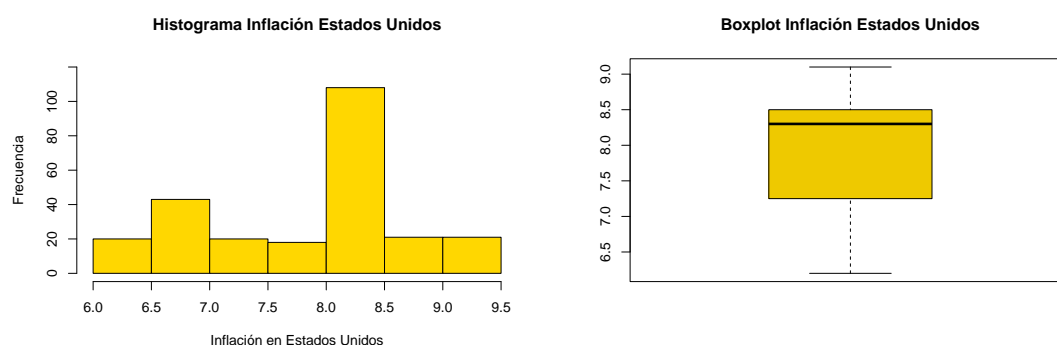


Figura 5: Histograma y Boxplot Inflación Estados Unidos

Con la información mostrada en las figuras anteriores podemos identificar a manera sospechosa que nuestros datos tienden a comportarse de una manera **Uniforme** con datos atípicos entre el 8.0% y 8.05% de inflación. A su vez, podemos detectar que los datos parecen demostrar una relación en sus frecuencias a niveles entre 0 y 20 ocurrencias en la mayoría de los niveles.

A continuación anexamos la distribución de datos a manera uniforme:

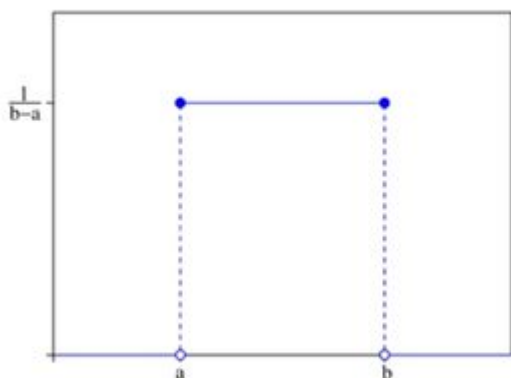


Figura 2.1: Distribución Uniforme

Variable Predictiva Dow Jones Industrial.

A continuación se muestra la tabla de frecuencias del DJI:

Lower	Upper	Main	Frequency	Percentage	CF	CPF
28000	29000	28500	1	0.4	1	0.4
29000	30000	29500	7	2.8	8	3.2
30000	31000	30500	17	6.8	25	10.0
31000	32000	31500	33	13.1	58	23.1
32000	33000	32500	37	14.7	95	37.8
33000	34000	33500	29	11.6	124	49.4
34000	35000	34500	55	21.9	179	71.3
35000	36000	35500	48	19.1	227	90.4
36000	37000	36500	24	9.6	251	100.0

De lo anterior podemos observar que el nivel más bajo se suitó a 28,000 puntos y el nivel máximo registrado se encontro a niveles de 37,000 puntos. Para ver estos datos de una manera más sencilla realizaremos un histograma y un gráfico de Boxplot.

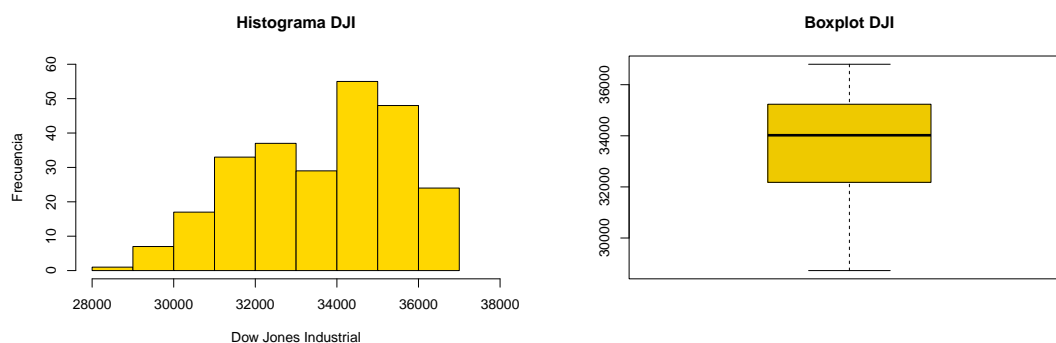


Figura 6: Histograma y Boxplot DJI

Con la información mostrada en las figuras anteriores podemos identificar a manera sospechosa que nuestros datos tienden a comportarse de una manera **Lognormal** la cual podemos ver su distribución en

la figura 2. A su vez, podemos detectar que los datos parecen demostrar una inclinación hacia la derecha del histograma, específicamente en niveles entre 32,000 y 35,000 puntos, lo cual lo podemos confirmar con nuestro Boxplot.

Variable Predictiva Precio Dólar.

A continuación se muestra la tabla de frecuencias del Precio del Dólar:

Lower	Upper	Main	Frequency	Percentage	CF	CPF
19.4	19.6	19.5	7	2.8	7	2.8
19.6	19.8	19.7	6	2.4	13	5.2
19.8	20.0	19.9	35	13.9	48	19.1
20.0	20.2	20.1	36	14.3	84	33.5
20.2	20.4	20.3	51	20.3	135	53.8
20.4	20.6	20.5	51	20.3	186	74.1
20.6	20.8	20.7	29	11.6	215	85.7
20.8	21.0	20.9	21	8.4	236	94.0
21.0	21.2	21.1	2	0.8	238	94.8
21.2	21.4	21.3	7	2.8	245	97.6
21.4	21.6	21.5	3	1.2	248	98.8
21.6	21.8	21.7	3	1.2	251	100.0

De lo anterior podemos observar que el nivel más bajo en el costo del dólar se encontró a 19.4 y el máximo a 21.8 pesos mexicanos. Para ver estos datos de una manera más sencilla realizaremos un histograma y un gráfico de Boxplot.

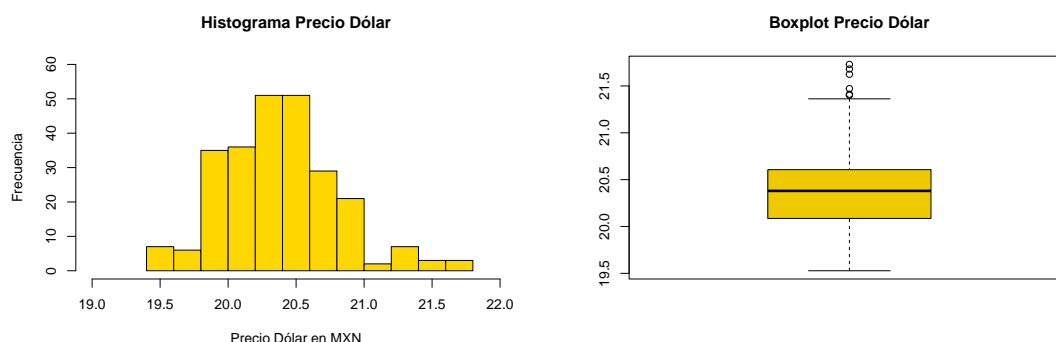


Figura 7: Histograma y Boxplot Precio Dólar

Con la información mostrada en las figuras anteriores podemos identificar a manera sospechosa que nuestros datos tienden a comportarse de una manera **Lognormal** la cual podemos ver su distribución en la figura 2. Complementado lo anterior, los datos tienden parecer que siguen una distribución normal pero por la naturaleza de los mismos (precios) y dado a que tiene caídas atípicas en sus frecuencias, optaremos seguir el supuesto que se distribuyen lognormal para posteriormente demostrarlo en el proyecto.

Análisis de Regresión Lineal Múltiple

Análisis Descriptivo y Multicolinealidad

Para realizar exitosamente una regresión lineal múltiple es necesario conocer primero los valores de correlación entre las variables y así poder plantear en sospecha que variables pueden realmente estar involucradas en el modelo.

Para ello, realizaremos nuestra matriz de correlación que se muestra a continuación:

	Apple	volumen_en_millones	SP500	inflacion_usa	DJI	precio_dolar
Apple	1.0000	0.0318	0.6446	-0.1354	-0.2489	0.3507
volumen_en_millones	0.0318	1.0000	-0.0344	-0.0019	-0.1484	0.1843
SP500	0.6446	-0.0344	1.0000	-0.6986	-0.6681	0.5431
inflacion_usa	-0.1354	-0.0019	-0.6986	1.0000	0.7775	-0.5120
DJI	-0.2489	-0.1484	-0.6681	0.7775	1.0000	-0.4116
precio_dolar	0.3507	0.1843	0.5431	-0.5120	-0.4116	1.0000

De lo anterior podemos concluir lo siguiente:

- Que la correlación entre Apple y el índice S&P 500 es considerable y positiva, lo cual nos indica que si sube el mercado, Apple tiende a también subir.
- En cambio al S&P 500, el índice Dow Jones tiene una correlación negativa respecto al precio de la acción de Apple, significando que si sube el precio en la cotización de la acción de Apple el índice tendrá una tendencia negativa.
- Que el precio del dólar tiende a bajar cuando sube la inflación en Estados Unidos.
- Que el volumen operado en la acción de Apple no está relacionada con el precio de su cotización, esto dado que la correlación entre ambas es de 0.031.
- Que la inflación en Estados Unidos. y el índice Dow Jones tienen una correlación fuerte y positiva, lo cual indica que un alza o baja repercute de casi igual forma en ambas.

Para reforzar lo planteado anteriormente, se agregará una matriz de dispersión y así confirmar los puntos establecidos.

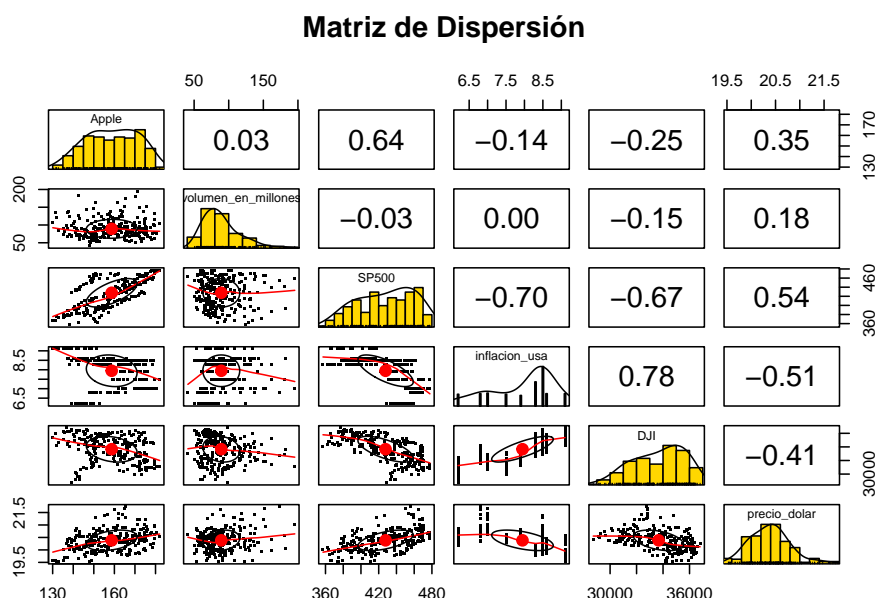


Figura 8: Matriz de Dispersión

De lo anterior podemos concluir lo siguiente:

- Que parece haber una relación lineal con la cotización de la acción de Apple y el índice S&P 500. Esto dado a que vemos una tendencia fuerte y positiva. Esto lo podemos confirmar también por su nivel de correlación el cual se encuentra en un nivel del 64%.

- Que la mayor parte del volumen operado se encontró en niveles entre 20.00 y 20.55 pesos mexicanos el dólar. Esto nos puede indicar que fue de más fácil acceso operar la acción en estos niveles del dólar.
- Que la cotización del S&P 500 tiene bajas considerables con la inflación de Estados Unidos a niveles superiores al 8.5%. Esto se debe principalmente a los fuertes movimientos del mercado, cada vez que sube la inflación tiene una repercusión negativa en los mercados bursátiles, en este caso en particular su correlación es de -70%.
- Que por el contrario al S&P 500 el índice Dow Jones se fortaleció a niveles de inflación mayores al 8.5%.
- A su vez, que el precio del dólar disminuyó considerablemente con una inflación mayor. Esto es normal en el mercado ya que los bancos centrales empiezan a subir la tasa de referencia, provocando que los inversionistas extranjeros tengan apetito por la adquisición de pesos mexicanos, disminuyendo la demanda del dólar y aumentando la del peso mexicano.

Al momento con la información obtenida, parece ser que todas nuestras variables impactan en el precio de la cotización de Apple, sin embargo podríamos decir que la única que no mantiene una posición fuerte dada su correlación es el volumen operado, no obstante, no se descartará dado que en la práctica si influye de manera significativa dado que si no hay corro de compra o venta se puede generar problema de oferta-demanda, ocasionando movimientos en su precio.

Dando seguimiento al modelo de regresión lineal múltiple, nos interesa revisar si el mismo tiene problemas de multicolinealidad. La multicolinealidad es la correlación alta entre 2 o más variables predictivas, es decir, que nuestras variables independientes (predictivas) no se correlacionen una con la otra.

A partir de este supuesto tomaremos los siguientes puntos para revisar si nuestro modelo presenta problemas de multicolinealidad:

$VIF = 1$ Significa que no existe correlación entre esta variable independiente y cualquier otra.

$1 < VIF < 5$ Sugiere una correlación moderada pero no sería necesario resolverla.

$VIF > 5$ Son niveles críticos de multicolinealidad.

Dicho esto, realizaremos la prueba para revisar que efectivamente nuestro modelo no presente problemas de multicolinealidad en sus datos.

La prueba nos arrojó los siguientes resultados:

	Volumen Operado	S&P 500	Inflación	DJI	Precio Dólar
VIF	1.155727	2.436469	3.216804	2.988866	1.61123

Tomando los criterios descritos y los valores obtenidos en la prueba podemos concluir que nuestro modelo **no corre con problemas de multicolinealidad**.

Pruebas de Significancia y Ajustes del Modelo Actual

Dada la información anterior, es necesario plantear nuestras pruebas de significancia del modelo de regresión lineal múltiple para medir la bondad de ajuste del modelo. Esta prueba determina si existe una relación lineal entre la variable respuesta Y y alguna de las variables regresoras X_1, X_2, X_3, X_4, X_5 . La hipótesis del modelo es la siguiente:

$$H_0 : B_1 = B_2 = \dots B_5 = 0$$

$$H_a : B_j \neq 0 \rightarrow j; 1, 2, \dots 5.$$

Al rechazar la hipótesis nula se concluye que al menos una de las variables regresoras contribuye significativamente al modelo. De no ser así, no podríamos afirmar que al menos una de nuestras variables influye en el modelo.

Dicho esto, la prueba nos arrojó los siguientes resultados:

```
##
## Call:
## lm(formula = data$Apple ~ data$volumen_en_millones + data$SP500 +
##     data$inflacion_usa + data$DJI + data$precio_dolar)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -15.403  -5.185   1.360   5.636  17.337
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -1.586e+02  3.319e+01  -4.779 3.04e-06 ***
## data$volumen_en_millones  2.131e-02  1.955e-02   1.090  0.2769
## data$SP500          4.088e-01  2.426e-02  16.852 < 2e-16 ***
## data$inflacion_usa    9.666e+00  1.040e+00   9.295 < 2e-16 ***
## data$DJI          -1.016e-04  4.426e-04  -0.230  0.8186
## data$precio_dolar    3.293e+00  1.496e+00   2.201  0.0287 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.667 on 245 degrees of freedom
## Multiple R-squared:  0.6215, Adjusted R-squared:  0.6138
## F-statistic: 80.47 on 5 and 245 DF,  p-value: < 2.2e-16
```

Ahora bien, para conocer si rechazamos nuestra hipótesis nula es necesario que nuestro *valor.p* sea menor que α . Por lo que desarrollamos lo siguiente:

$$\begin{aligned} \text{valor.p} &< \alpha \\ 2.2e^{-16} &< 0.05 \end{aligned}$$

Dado a que efectivamente nuestro *valor.p* es menor que α podemos afirmar con un nivel del 95% de confianza que **al menos una de las variables regresoras contribuye significativamente al modelo**. Por otra parte, podemos destacar que nuestro modelo aparte de ser significativo cumple con un desempeño del **61.38%**.

Continuando con el análisis es necesario conocer y plantear la ecuación del modelo. Para ello tomaremos los datos de la prueba realizada en el punto anterior. La ecuación planteada para el modelo que incluye todas las variables es la siguiente:

$$\hat{y} = 0.02131X_1 + 0.4088X_2 + 9.666X_3 - 0.0001016X_4 + 3.293X_5 - 158.6$$

Dada la ecuación del modelo podemos decir que todas las variables influyen positivamente en nuestra variable de respuesta a excepción de la variable del índice Dow Jones la cual se ve con una repercusión negativa sobre la función. Esto también hace concordancia con la matriz de correlación de nuestros datos.

Regresiones Posibles y Mejor Modelo

Anteriormente pudimos observar que nuestro modelo que contempla todas nuestras variables predictivas cumple con la significancia de que al menos una de las variables influye en nuestro modelo y tuvo un desempeño por encima del 60%, sin embargo no podemos afirmar que este sea el mejor modelo a trabajar, es por ello que se realizarán las pruebas para todos los modelos con 4, 3, 2 y 1 variable para conocer puntualmente cuál de ellos presentan un mejor desempeño y una mayor significancia dentro del mismo.

Para este análisis se contemplaron las variables predictivas en el siguiente orden; Volumen Operado(1), S&P 500(2), Inflación en Estados Unidos(3), Dow Jones(4) y el Precio del Dólar(5). Las pruebas nos arrojaron los siguientes resultados:

No.	1	2	3	4	5	Desempeño	valor_p
1	x	x	x	x	-	0.607800	2.2e-16
2	x	x	x	-	x	0.615300	2.2e-16
3	x	x	-	x	x	0.479700	2.2e-16
4	x	-	x	x	x	0.169500	3.852e-10
5	-	x	x	x	x	0.613500	2.2e-16
6	x	x	x	-	-	0.609400	2.2e-16
7	x	x	-	x	-	0.481800	2.2e-16
8	x	x	-	-	x	0.411500	2.2e-16
9	x	-	x	x	-	0.059670	0.000397
10	x	-	x	-	x	0.116600	2.33e-07
11	x	-	-	x	x	0.127500	5.201e-08
12	-	x	x	x	-	0.604600	2.2e-16
13	-	x	x	-	x	0.614600	2.2e-16
14	-	x	-	x	x	0.469200	2.2e-16
15	-	-	x	x	x	0.164100	2.936e-10
16	x	x	-	-	-	0.413700	2.2e-16
17	x	-	x	-	-	0.011430	0.08878
18	x	-	-	x	-	0.054400	0.0003592
19	x	-	-	-	x	0.117100	7.302e-08
20	-	x	x	-	-	0.606000	2.2e-16
21	-	x	-	x	-	0.470900	2.2e-16
22	-	x	-	-	x	0.410800	2.2e-16
23	-	-	x	x	-	0.062960	0.0001163
24	-	-	x	-	x	0.118600	5.884e-08
25	-	-	-	x	x	0.129200	1.313e-08
26	x	-	-	-	-	-0.003002	0.6162
27	-	x	-	-	-	0.413200	2.2e-16
28	-	-	x	-	-	0.014400	0.03195
29	-	-	-	x	-	0.058170	6.72e-05
30	-	-	-	-	x	0.119500	1.123e-08

De lo anterior pudimos observar que el modelo que presenta un mejor desempeño es el que incluye las variables predictivas del **Volumen Operado, S&P 500, Inflación en Estados Unidos y el Precio del Dólar**. La variable predictiva del Dow Jones **será retirada del modelo** para así tener un mejor desempeño en el modelo de regresión lineal múltiple. Este nuevo modelo presenta un desempeño del **61.53%** vs **61.38%** por lo que pretende tener una mejor estimación.

Dicho esto, se realizará de nuevo las pruebas de significancia y el ajuste del nuevo modelo a utilizar.

Prueba de Significancia y Ajuste del Modelo 2

La prueba de significancia del nuevo modelo de regresión lineal múltiple sirve para medir la bondad de ajuste del mismo. Esta prueba determina si existe una relación lineal entre la variable respuesta Y y alguna de las variables regresoras X_1, X_2, X_3, X_5 . La hipótesis del modelo es la siguiente:

$$H_0 : B_1 = B_2 = B_3 = B_5 = 0$$

$$H_a : B_j \neq 0 \rightarrow j; 1, 2, 3, 5.$$

Al rechazar la hipótesis nula se concluye que al menos una de las variables regresoras contribuye significativamente al modelo.

Dicho esto, la prueba nos arrojó los siguientes resultados:

```
##
## Call:
```

```
## lm(formula = data$Apple ~ data$volumen_en_millones + data$SP500 +
##      data$inflacion_usa + data$precio_dolar)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -15.381   -5.277    1.289    5.633   17.382
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -160.82063    31.68206  -5.076 7.61e-07 ***
## data$volumen_en_millones    0.02258     0.01872   1.206  0.2291
## data$SP500         0.41064     0.02291  17.922 < 2e-16 ***
## data$inflacion_usa     9.52203     0.82929  11.482 < 2e-16 ***
## data$precio_dolar     3.24684     1.47968   2.194  0.0292 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.653 on 246 degrees of freedom
## Multiple R-squared:  0.6215, Adjusted R-squared:  0.6153
## F-statistic: 101 on 4 and 246 DF, p-value: < 2.2e-16
```

Al igual que en modelo pasado, necesitamos conocer si rechazamos nuestra hipótesis nula por lo que es necesario que nuestro *valor.p* sea menor que α . Por lo que desarrollamos lo siguiente:

$$\begin{aligned} \text{valor.p} &< \alpha \\ 2.2e^{-16} &< 0.05 \end{aligned}$$

Dado a que efectivamente nuestro *valor.p* es menor que α podemos afirmar con un nivel del 95% de confianza que **al menos una de las variables regresoras contribuye significativamente al modelo**. Por otra parte, podemos destacar que nuestro modelo aparte de ser significativo cumple con un desempeño del **61.53%**.

A su vez es necesario conocer y plantear la ecuación del modelo. Para ello tomaremos los datos de la prueba realizada en el punto anterior. La ecuación planteada para el modelo final que incluye las variables de **Volumen Operado, S&P 500, Inflación en Estados Unidos y el Precio del Dólar** es la siguiente:

$$\hat{y} = 0.02258X_1 + 0.41064X_2 + 9.52203X_3 + 3.24684X_5 - 160.82063$$

Con la nueva ecuación del modelo podemos concluir que todas nuestras variables impactan de manera positiva a nuestro modelo, en cambio al modelo anterior, el Dow Jones impactaba de manera negativa. Por otro lado vemos que la inflación en Estados Unidos toma un gran papel en el precio de la acción de Apple como también el precio del dólar. Llevado a la práctica esto puede resultar fiable dado que la inflación impacta directamente a los mercados bursátiles, esto principalmente porque el poder adquisitivo del mercado disminuye y buscan mejores alternativas para conservar el valor de su dinero.

Intervalos de Confianza del Modelo 2

Ya que tenemos nuestra ecuación del modelo es importante revisar si este presenta pruebas de tener una **regresión al origen**.

Para ello es necesario conocer nuestros intervalos de confianza los cuales se muestran a continuación:

	Intercept	Volumen Operado	S&P 500	Inflación	Precio Dólar
2.5%	-223.22333	-0.0143025	0.3655118	7.888617	0.3323884
97.5%	-98.41794	0.0594556	0.4557714	11.155436	6.1612857

De lo anterior podemos concluir que nuestro modelo **no presenta evidencia alguna de tener regresión al origen**, esto dado que nuestros valores van desde -223.22 al -98.41 por lo cual no tiene un punto donde pase por el 0.

Supuestos de los Residuos

Para validar que nuestro modelo realmente es funcional es importante revisar el cumplimiento de los supuestos de los residuos. Estos constan de 3 puntos:

- El término del error tiene varianza constante.
- El término del error tiene *media* = 0 y cumplen con una distribución normal
- Los errores son independientes

Revisando los puntos anteriores se realizaron las siguientes pruebas:

Prueba de Normalidad y Media = 0

Para ver el cumplimiento de este supuesto donde buscamos que el termino del error sigan una distribución normal y que su media sea igual a 0 es necesario plantear la siguiente prueba de hipótesis:

$$H_0 : \text{Residuales.Distribucion.Normal/Media} = 0$$

$$H_a : \text{Los.datos.proviene.n.de.otra.distribución}$$

La prueba arrojó los siguientes resultados:

```
##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  res
## D = 0.083013, p-value = 0.0629
## alternative hypothesis: two-sided
```

Para interpretar esta prueba y poder llegar a una conclusión tenemos que comprobar los niveles de nuestro *valor.p* contra nuestro valor de α .

$$\text{valor.p} < \alpha$$

$$0.0629 > 0.05$$

Dado que nuestro *valor.p* es mayor que nuestro valor de α **no rechazamos H0** y podemos afirmar con un nivel del 95% de confianza que **los residuales provienen de una distribución normal y tienen media igual a 0**.

Prueba de Varianza Constante

Para buscar que el termino del error tenga varianza constante necesitamos analizar el gráfico de dispersión de los valores ajustados vs los residuales. Para ello se anexa la siguiente figura:

-
-
-
-
-

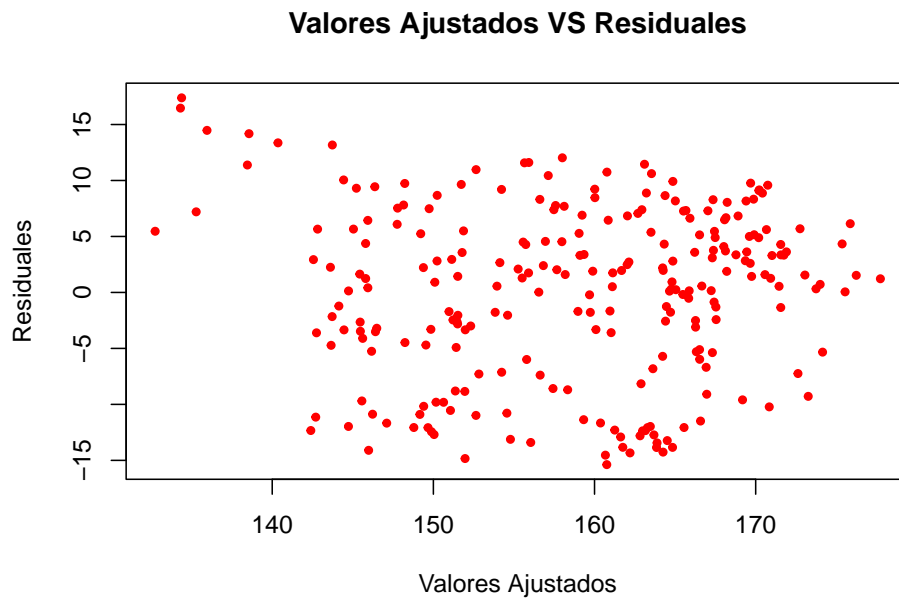


Figura 9: Dispersión de Valores Ajustados VS Residuales

•
•

Analizando el gráfico anterior podemos observar que los datos no presentan ningún patrón en la dispersión de los mismos, por lo tanto, **podemos afirmar que la regresión sigue una varianza constante.**

Supesto de Independencia

Para finalizar el análisis del cumplimiento de los supuestos de los residuos es necesario comprobar que los errores son independientes y no se relacionan unos con los otros. Para ello tenemos que plantear la siguiente prueba de hipótesis:

$$H_0 : \text{Residuales.Independientes}$$

$$H_a : \text{Residuales.Dependientes}$$

La prueba arrojó los siguientes resultados:

```
##
## Box-Pierce test
##
## data:  res
## X-squared = 234.1, df = 1, p-value < 2.2e-16
```

Para interpretar esta prueba y poder llegar a una conclusión tenemos que comprobar los niveles de nuestro *valor.p* contra nuestro valor de α .

$$\begin{aligned} \text{valor.p} &< \alpha \\ 2.2e^{-16} &< 0.05 \end{aligned}$$

Dado que nuestro *valor.p* es menor que nuestro valor de α **rechazamos H_0** y podemos afirmar con un nivel del 95% de confianza que **los residuales son dependientes.**

Dado que uno de nuestros supuestos resultó ser rechazado, tenemos que aplicar una transformación a nuestros datos.

Corrección de la Autocorrelación de 1er Orden Cochrane Orcutt

Si se comprueba a través del estadístico de Durbin-Watson que el término de error está serialmente correlacionado con el tiempo, entonces la inferencia estadística estándar tal como se aplica normalmente a regresiones no es válida porque los errores estándar se estimaron con sesgo. Para evitar este problema, los residuos deben ser modelados. Si se encuentra que el proceso de generación de los residuos es un proceso estacionario de primer orden con estructura autorregresiva, con los errores siendo ruido blanco, el procedimiento de Cochrane-Orcutt puede ser utilizado para transformar el modelo

Dando seguimiento, se realizará primero la prueba de hipótesis a través del estadístico de Durbin-Watson para comprobar la autocorrelación de nuestros datos, la cual se muestra a continuación:

$$H_0 : \text{Independencia}$$

$$H_a : \text{Autocorrelación}$$

La prueba Durbin-Watson nos arrojó los siguientes resultados:

```
##
## Durbin-Watson test
##
## data: reg2
## DW = 0.065532, p-value < 2.2e-16
## alternative hypothesis: true autocorrelation is not 0
```

Para interpretar esta prueba y poder llegar a una conclusión tenemos que comprobar los niveles de nuestro *valor.p* contra nuestro valor de α .

$$\text{valor.p} < \alpha$$

$$2.2e^{-16} < 0.05$$

Dado a que **rechazamos H_0** podemos concluir que con un nivel del 95% de confianza que nuestros datos tienen autocorrelación y podremos aplicar el método de la corrección de la autocorrelación de 1er orden de Cochrane Orcutt.

Ahora bien para realizar esta prueba es necesario plantear nuestra prueba de hipótesis:

$$H_0 : \text{Independencia}$$

$$H_a : \text{Autocorrelación}$$

El modelo de corrección de Cochrane Orcutt nos arrojó los siguientes resultados:

```
## Call:
## lm(formula = data$Apple ~ data$volumen_en_millones + data$SP500 +
##     data$inflacion_usa + data$precio_dolar)
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -49.8332497   22.0754911   -2.257  0.02486 *
## data$volumen_en_millones  0.0100764   0.0049095    2.052  0.04119 *
## data$SP500         0.4704695   0.0186769   25.190 < 2e-16 ***
## data$inflacion_usa  0.6144166   1.2161306    0.505  0.61386
## data$precio_dolar   0.5018506   0.8101773    0.619  0.53621
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.6695 on 245 degrees of freedom
## Multiple R-squared:  0.7245 , Adjusted R-squared:  0.72
```

```
## F-statistic: 161.1 on 4 and 245 DF, p-value: < 2.304e-67
##
## Durbin-Watson statistic
## (original): 0.06553 , p-value: 2.686e-55
## (transformed): 1.81323 , p-value: 7.05e-02
```

Para interpretar esta prueba y poder llegar a una conclusión tenemos que comprobar los niveles de nuestro *valor.p* contra nuestro valor de α .

$$\begin{aligned} \text{valor.p} &< \alpha \\ 0.0705 &> 0.05 \end{aligned}$$

Dado a que **no rechazamos H_0** podemos concluir que con un nivel del 95% de confianza que nuestros residuos son independientes.

Prueba de Significancia y Ajuste del Modelo Final

Esta prueba determina si existe una relación lineal entre la variable respuesta Y y alguna de las variables regresoras X_1, X_2, X_3, X_5 con el nuevo modelo corregido. La hipótesis del modelo es la siguiente:

$$\begin{aligned} H_0 : B_1 = B_2 = B_3 = B_5 = 0 \\ H_a : B_j \neq 0 \rightarrow j; 1, 2, 3, 5. \end{aligned}$$

Al rechazar la hipótesis nula se concluye que al menos una de las variables regresoras contribuye significativamente al modelo.

Dicho esto, la prueba nos arrojó los siguientes resultados:

```
## Call:
## lm(formula = data$Apple ~ data$volumen_en_millones + data$SP500 +
##     data$inflacion_usa + data$precio_dolar)
##
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -49.8332497   22.0754911   -2.257  0.02486 *
## data$volumen_en_millones  0.0100764   0.0049095    2.052  0.04119 *
## data$SP500         0.4704695   0.0186769   25.190 < 2e-16 ***
## data$inflacion_usa    0.6144166   1.2161306    0.505  0.61386
## data$precio_dolar     0.5018506   0.8101773    0.619  0.53621
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.6695 on 245 degrees of freedom
## Multiple R-squared:  0.7245 , Adjusted R-squared:  0.72
## F-statistic: 161.1 on 4 and 245 DF, p-value: < 2.304e-67
##
## Durbin-Watson statistic
## (original): 0.06553 , p-value: 2.686e-55
## (transformed): 1.81323 , p-value: 7.05e-02
```

Al igual que en modelo pasado, necesitamos conocer si rechazamos nuestra hipótesis nula por lo que es necesario que nuestro *valor.p* sea menor que α . Por lo que desarrollamos lo siguiente:

$$\begin{aligned} \text{valor.p} &< \alpha \\ 2.304e^{-67} &< 0.05 \end{aligned}$$

Dado a que efectivamente nuestro *valor.p* es menor que α podemos afirmar con un nivel del 95% de confianza que **al menos una de las variables regresoras contribuye significativamente al modelo**. Por otra parte, podemos destacar que nuestro modelo aparte de ser significativo cumple con un desempeño del **72%**.

A su vez es necesario conocer y plantear la ecuación del modelo. Para ello tomaremos los datos de la prueba realizada en el punto anterior. La ecuación planteada para el modelo final que incluye las variables de **Volumen Operado, S&P 500, Inflación en Estados Unidos y el Precio del Dólar** es la siguiente:

$$\hat{y} = 0.0100764X_1 + 0.4704695X_2 + 0.6144166X_3 + 0.5018506X_5 - 49.8332497$$

Con la nueva ecuación del modelo podemos concluir que todas nuestras variables impactan de manera positiva a nuestro modelo, en cambio al modelo anterior, vemos que los valores de nuestras B_j de las variables predictivas disminuyeron, esto se debe precisamente a la corrección aplicada.

Intervalos de Confianza del Modelo Final

Ya que tenemos nuestra ecuación del modelo es importante revisar si este presenta pruebas de tener una **regresión al origen**.

Para ello es necesario conocer nuestros intervalos de confianza los cuales se muestran a continuación:

	Intercept	Volumen Operado	S&P 500	Inflación	Precio Dólar
2.5%	-93.315210	0.0004062	0.4336818	-1.780988	-1.093951
97.5%	-6.351289	0.0197466	0.5072573	3.009822	2.097652

De lo anterior podemos concluir que nuestro modelo **no presenta evidencia alguna de tener regresión al origen**, esto dado que nuestros valores van desde -93.3152 al -6.3512 por lo cual no tiene un punto donde pase por el 0.

Pruebas de Bondad de Ajuste

Una prueba de bondad de ajuste permite rectificar la hipótesis de que una variable aleatoria sigue cierta distribución de probabilidad y se utiliza en situaciones donde se requiere comparar una distribución observada con una teórica o hipotética, compararla con datos históricos o con la distribución conocida de otra población.

Es por ello que realizaremos una prueba de bondad de ajuste para cada una de las variables a utilizar en nuestra ecuación final del modelo.

Bondad de Ajuste para el Precio de la Acción de Apple

Para esta variable se planteó en el análisis exploratorio de datos a manera de sospecha que los datos siguen una distribución lognormal. A continuación se muestra nuevamente el histograma del precio de la acción de Apple.

-
-
-
-
-
-
-
-
-

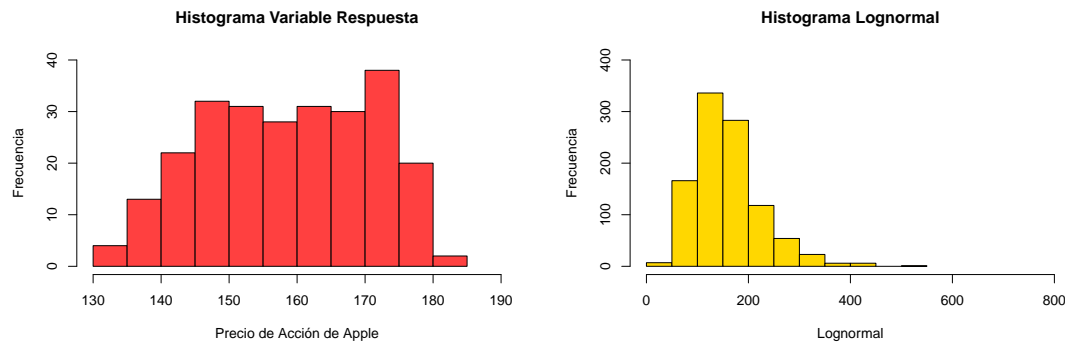


Figura 10: Histograma Precio Acción de Apple y Distribución Lognormal

Ahora bien, es necesario realizar la estimación de nuestros parámetros para plantear de manera correcta nuestra prueba de hipótesis. Esta prueba se realizará apartir de la **estimación por máxima verosimilitud (MLE)**.

```
## Fitting of the distribution 'lnorm' by maximum likelihood
## Parameters:
##      estimate Std. Error
## meanlog 5.06391755      NA
## sdlog   0.07844668      NA
```

De lo anterior podemos observar que nuestras estimaciones de los parámetros fueron de $\mu = 5.0639$ y $\sigma = 0.07844$. Dicho esto, podemos plantear de manera correcta nuestra hipótesis correspondiente:

$$H_0 : \text{Distribución.Lognormal} | \mu = 5.0639 \wedge \sigma = 0.07844$$

$$H_a : \text{Otra.Distribución}$$

Planteada nuestra hipótesis realizaremos nuestra prueba *Kolmogorov Smirnov*:

```
## Warning in ks.test.default(data$Apple, plnorm, mean = 5.0639, sd = 0.07844):
## ties should not be present for the Kolmogorov-Smirnov test

##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data:  data$Apple
## D = 0.072356, p-value = 0.1444
## alternative hypothesis: two-sided
```

Por lo tanto, desarrollamos nuestra hipótesis:

$$\text{valor.p} < \alpha$$

$$0.1444 > 0.05$$

Dado que **no rechazamos** H_0 podemos concluir con un nivel del 95% de confianza que nuestros datos **proviene de una distribución lognormal** con parámetros de $\mu = 5.0639$ y $\sigma = 0.07844$.

Bondad de Ajuste para el Volumen Operado de la Acción de Apple

Para esta variable se planteó en el análisis exploratorio de datos a manera de sospecha que los datos siguen una distribución lognormal. A continuación se muestra nuevamente el histograma del volumen operado en millones de dólares.

•

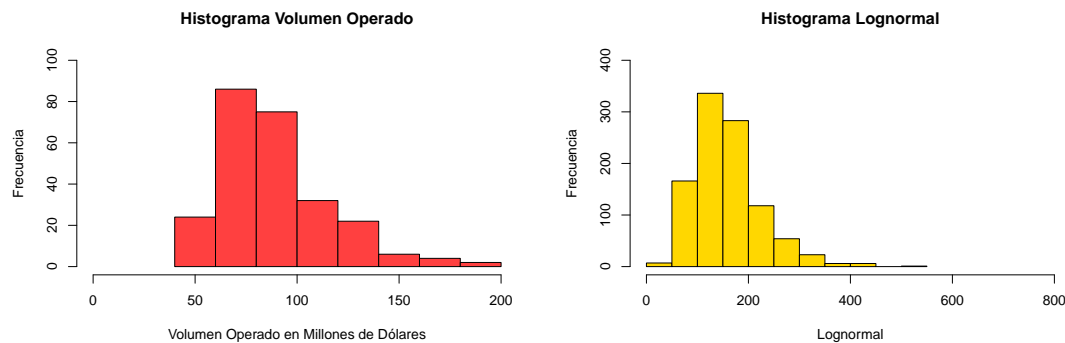


Figura 11: Histograma Volumen Operado y Distribución Lognormal

Ahora bien, es necesario realizar la estimación de nuestros parámetros para plantear de manera correcta nuestra prueba de hipótesis. Esta prueba se realizará apartir de la **estimación por máxima verosimilitud (MLE)**.

```
## Fitting of the distribution 'lnorm' by maximum likelihood
## Parameters:
##      estimate Std. Error
## meanlog 4.4513909      NA
## sdlog    0.2781838      NA
```

De lo anterior podemos observar que nuestras estimaciones de los parámetros fueron de $\mu = 4.45139$ y $\sigma = 0.278183$. Dicho esto, podemos plantear de manera correcta nuestra hipótesis correspondiente:

$$H_0 : \text{Distribución.Lognormal} | \mu = 4.45139 \wedge \sigma = 0.2781$$

$$H_a : \text{Otra.Distribución}$$

Planteada nuestra hipótesis realizaremos nuestra prueba *Kolmogorov Smirnov*:

```
##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data: data$volumen_en_millones
## D = 0.056568, p-value = 0.398
## alternative hypothesis: two-sided
```

Por lo tanto, desarrollamos nuestra hipótesis:

$$\text{valor.p} < \alpha$$

$$0.398 > 0.05$$

Dado que **no rechazamos** H_0 podemos concluir con un nivel del 95% de confianza que nuestros datos **proviene de una distribución lognormal** con parámetros de $\mu = 0.398$ y $\sigma = 0.2781$.

Bondad de Ajuste para el S&P 500

Para esta variable se planteó en el análisis exploratorio de datos a manera de sospecha que los datos siguen una distribución lognormal. A continuación se muestra nuevamente el histograma del S&P 500.

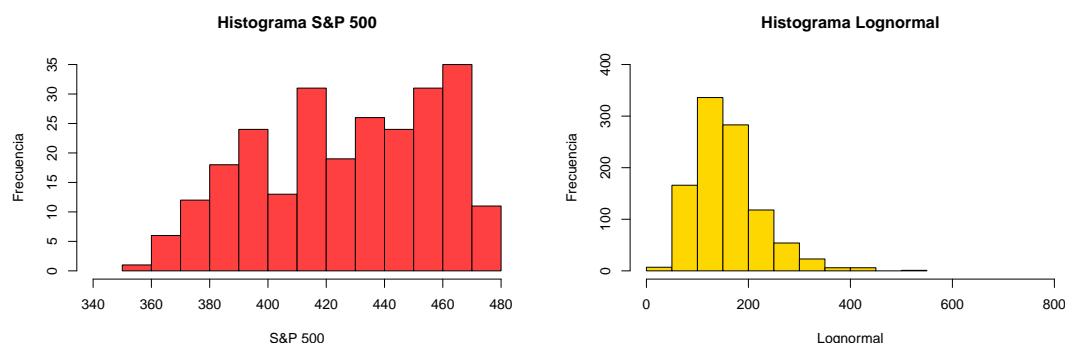


Figura 12: Histograma SP 500 y Distribución Lognormal

Ahora bien, es necesario realizar la estimación de nuestros parámetros para plantear de manera correcta nuestra prueba de hipótesis. Esta prueba se realizará apartir de la **estimación por máxima verosimilitud (MLE)**.

```
## Fitting of the distribution 'lnorm' by maximum likelihood
## Parameters:
##      estimate Std. Error
## meanlog 6.0558367      NA
## sdlog   0.0738248      NA
```

De lo anterior podemos observar que nuestras estimaciones de los parámetros fueron de $\mu = 6.05583$ y $\sigma = 0.073824$. Dicho esto, podemos plantear de manera correcta nuestra hipótesis correspondiente:

$$H_0 : \text{Distribución.Lognormal} | \mu = 6.05583 \wedge \sigma = 0.073824$$

$$H_a : \text{Otra.Distribución}$$

Planteada nuestra hipótesis realizaremos nuestra prueba *Kolmogorov Smirnov*:

```
## Warning in ks.test.default(data$SP500, plnorm, mean = 6.0558367, sd =
## 0.0738248): ties should not be present for the Kolmogorov-Smirnov test

##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data: data$SP500
## D = 0.088543, p-value = 0.03907
## alternative hypothesis: two-sided
```

Por lo tanto, desarrollamos nuestra hipótesis:

$$\text{valor.p} < \alpha$$

$$0.03907 < 0.05$$

Dado que **rechazamos** H_0 podemos concluir con un nivel del 95% de confianza que nuestros datos **proviene de alguna otra distribución**.

Bondad de Ajuste para la Inflación en Estados Unidos

Para esta variable se planteó en el análisis exploratorio de datos a manera de sospecha que los datos siguen una distribución uniforme. A continuación se muestra nuevamente el histograma de la inflación en Estados Unidos.

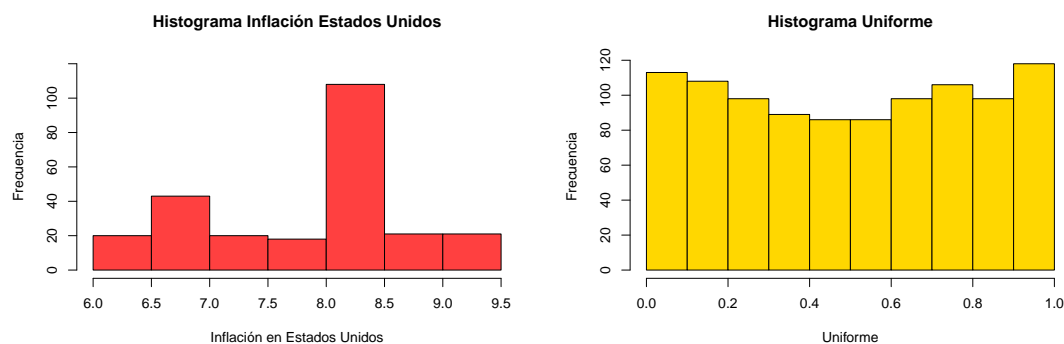


Figura 13: Histograma Inflación en Estados Unidos y Distribución Uniforme

Dado que observamos que se presentan datos atípicos en los valores entre el 8.0 y el 8.5 de inflación en Estados Unidos, se optará por quitar estos valores y trabajar sobre un nuevo panorama para analizar su comportamiento de manera correcta.

Quitando estos datos atípicos podemos observar un planteamiento:

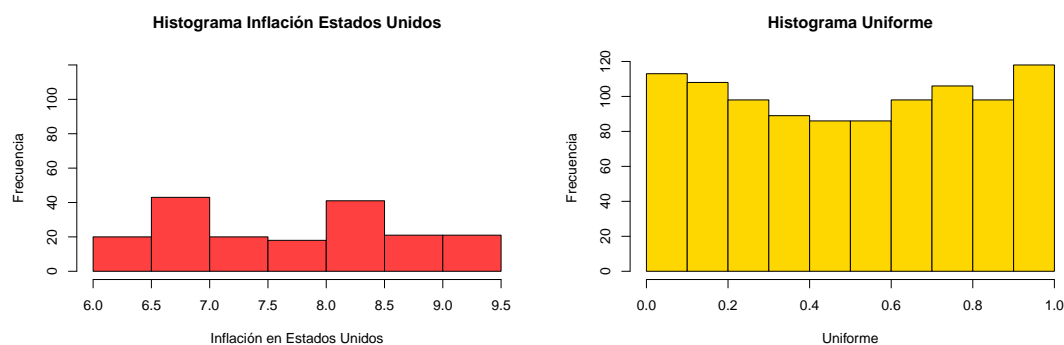


Figura 14: Histograma Inflación en Estados Unidos y Distribución Uniforme

Ahora bien, es necesario realizar la estimación de nuestros parámetros para plantear de manera correcta nuestra prueba de hipótesis. Esta prueba se realizará apartir de la **estimación por máxima verosimilitud (MLE)**.

```
## Fitting of the distribution 'unif' by maximum likelihood
## Parameters:
##      estimate Std. Error
## min      6.2         NA
## max      9.1         NA
```

De lo anterior podemos observar que nuestras estimaciones de los parámetros fueron de $min = 6.2$ y $max = 9.1$. Dicho esto, podemos plantear de manera correcta nuestra hipótesis correspondiente:

$$H_0 : \text{Distribución.Uniforme} | min = 6.2 \wedge max = 9.1$$

$$H_a : \text{Otra.Distribución}$$

Planteada nuestra hipótesis realizaremos nuestra prueba *Kolmogrov Smirnov*:

```
## Warning in ks.test.default(inflacion$inflacion_usa, punif, min = 6.2, max =
## 9.1): ties should not be present for the Kolmogorov-Smirnov test
```

```
##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data: inflacion$inflacion_usa
## D = 0.17522, p-value = 2.477e-05
## alternative hypothesis: two-sided
```

Por lo tanto, desarrollamos nuestra hipótesis:

$$\begin{aligned} \text{valor}.p &< \alpha \\ 2.477e^{-5} &< 0.05 \end{aligned}$$

Dado que **rechazamos** H_0 podemos concluir con un nivel del 95% de confianza que nuestros datos **proviene de alguna otra distribución**.

Bondad de Ajuste para el Precio del Dólar

Para esta variable se planteó en el análisis exploratorio de datos a manera de sospecha que los datos siguen una distribución lognormal. A continuación se muestra nuevamente el histograma del precio del dólar.

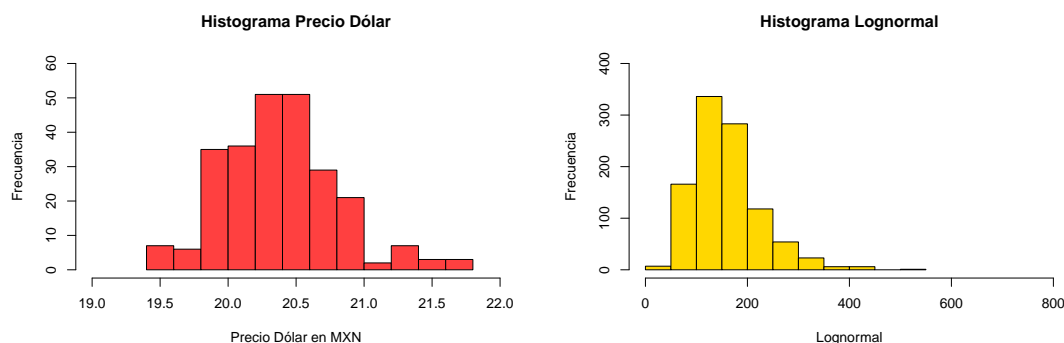


Figura 15: Histograma Precio Dólar y Distribución Lognormal

Ahora bien, es necesario realizar la estimación de nuestros parámetros para plantear de manera correcta nuestra prueba de hipótesis. Esta prueba se realizará apartir de la **estimación por máxima verosimilitud (MLE)**.

```
## Fitting of the distribution 'lnorm' by maximum likelihood
## Parameters:
##          estimate Std. Error
## meanlog 3.01454184          NA
## sdlog    0.02005581          NA
```

De lo anterior podemos observar que nuestras estimaciones de los parámetros fueron de $\mu = 3.0145418$ y $\sigma = 0.0200558$. Dicho esto, podemos plantear de manera correcta nuestra hipótesis correspondiente:

$$\begin{aligned} H_0 &: \text{Distribución.Lognormal} | \mu = 3.0145418 \wedge \sigma = 0.0200558 \\ H_a &: \text{Otra.Distribución} \end{aligned}$$

Planteada nuestra hipótesis realizaremos nuestra prueba *Kolmogrov Smirnov*:


```
## Warning in ks.test.default(data$precio_dolar, plnorm, mean = 3.0145418, : ties
## should not be present for the Kolmogorov-Smirnov test

##
## Asymptotic one-sample Kolmogorov-Smirnov test
##
## data: data$precio_dolar
## D = 0.057225, p-value = 0.3837
## alternative hypothesis: two-sided
```

Por lo tanto, desarrollamos nuestra hipótesis:

$$\begin{aligned} \text{valor.p} &< \alpha \\ 0.3837 &> 0.05 \end{aligned}$$

Dado que **no rechazamos** H_0 podemos concluir con un nivel del 95% de confianza que nuestros datos **proviene de una distribución lognormal** con parámetros de $\mu = 3.0145418$ y $\sigma = 0.020058$.

Prueba de Hipotesis sobre los Parámetros

Para llevar a cabo estas pruebas únicamente nos enfocaremos en las siguientes variables:

- Precio de la Acción de Apple
- Precio del Dólar

De primer instancia y de acuerdo a nuestros datos, se estipuló que en 251 días la acción de Apple presentó ganancias en 125 días, por otro lado el precio del dólar solo en 115 días. ¿Existe diferencia significativa entre la proporción de las ganancias de ambas cotizaciones?

Planteando la prueba de hipótesis obtenemos lo siguiente:

$$\begin{aligned} H_0 &: \text{Existe la misma proporción de ganancias} \\ H_a &: \text{Existe diferente proporción de ganancias} \end{aligned}$$

La prueba nos arrojó los siguientes resultados:

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data: c(125, 115) out of c(251, 251)
## X-squared = 0.64666, df = 1, p-value = 0.4213
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.05146733 0.13114861
## sample estimates:
## prop 1 prop 2
## 0.4980080 0.4581673
```

Por lo tanto, desarrollamos nuestra hipótesis:

$$\begin{aligned} \text{valor.p} &< \alpha \\ 0.4213 &> 0.05 \end{aligned}$$

Dado que **no rechazamos** H_0 podemos concluir con un nivel del 95% de confianza que **existe la misma proporción de ganancias entre ambas cotizaciones**.

Conclusiones

En el análisis exploratorio de los datos se fue planteando a manera de sospecha a que tipo de distribución se asimilaban los datos, sin embargo, más adelante se pudo afirmar que solo 3 de las 5 variables del modelo final cumplían con las aseveraciones planteadas.

Por otro lado, en el ajuste del modelo se percató que el supuesto de independencia de los residuales no se cumplía, por lo que se tuvo que recurrir a la corrección de autocorrelación por el método de Cochrane Orcutt. Gracias a este procedimiento se estableció un último modelo el cual incluye a las variables de **Volumen Operado, S&P 500, Inflación en Estados Unidos y el Precio del Dólar**, a continuación se muestra la ecuación del mismo:

$$\hat{y} = 0.0100764X_1 + 0.4704695X_2 + 0.6144166X_3 + 0.5018506X_5 - 49.8332497$$

Este último modelo presentó un desempeño del 72% y las variables significativas fueron el volumen operado y el S&P 500. De la ecuación anterior se puede responder las incógnitas planteadas al principio del estudio.

- Si aumenta el S&P 500, ¿cómo se ve afecta la acción de Apple?
 - Se puede ver afectada de manera positiva dado a que existe una gran correlación entre ambas variables.
- En un escenario de alta inflación, ¿la acción de Apple sufre decrementos?
 - Si, dado a que entre ambas variables existe una correlación negativa y a su vez, esto se debe principalmente a los fuertes movimientos del mercado, cada vez que sube la inflación tiene una repercusión negativa en los mercados bursátiles. También los altos niveles de inflación impacta al S&P 500 con niveles hasta del -70% por lo cual nuestra acción de Apple también se ve afectada.
- Con una cotización “cara” del precio del dólar, ¿aumenta el precio de la acción de Apple?
 - Para esta pregunta podríamos concluir que si, dado que ambas variables siguen una tendencia positiva y con un nivel de correlación de 35%. Esto no significa que siempre que suba el precio del dólar va a subir el precio de la acción de Apple, principalmente porque no se demostró que esta variable impacta de manera significativa en nuestro modelo.

Para finalizar, se hará una prueba para ver el desempeño de nuestro modelo en un caso práctico. Para ello se recolectaron los datos del día 05 de Octubre del 2022 para hacerle una prueba al modelo y ver que tan lejano o cercano se encuentra la estimación al dato real en ese día.

Los precios y cotizaciones de nuestras variables regresoras al día 05 de Octubre de 2022 fueron las siguientes:

Fecha	Volumen Operado	SP500	Inflacion	Dólar
2022-10-05	79.47	375	7.7	20.04

Aplicando esta información en nuestra ecuación del modelo obtenemos lo siguiente:

$$\hat{y} = 0.0100764(79.47) + 0.4704695(375) + 0.6144166(7.7) + 0.5018506(20.04) - 49.8332497$$

$$\hat{y} = 0.800771508 + 176.4260625 + 4.73100782 + 10.05708602 - 49.8332497$$

$$\hat{y} = 142.1816$$

De lo anterior podemos observar que el modelo arrojó que el día 05 de Octubre de 2022 se estima que el precio de la acción de Apple se encuentre en 142.18 dólares. A continuación se muestra una imagen donde podemos ver los niveles de la acción de Apple durante ese día.



Figura 16: Niveles de la Acción de Apple 05-Oct-22

Como conclusión podemos observar en el recuadro verde de la Figura 16 que el precio en la acción de Apple durante el día 05 de Octubre de 2022 se encontró en niveles entre 142.9 y 147.4 dólares. Tomando en cuenta que nuestro modelo estimó 142.18 dólares en ese día podemos concluir que nuestra estimación es muy cercana a la ocurrencia real.

Bibliografías

Apple Inc. (AAPL). (s/f). Yahoo.com. Recuperado el 5 de diciembre de 2022, de <https://finance.yahoo.com/quote/AAPL/history?p=AAPL>

Dow Jones Industrial Average (^DJI). (s/f). Yahoo.com. Recuperado el 5 de diciembre de 2022, de <https://finance.yahoo.com/quote/%5EDJI/history?p=%5EDJI>

Pinchas Cohen/Investing.com, & Picerno, J. (s/f). EU - Inflación - interanual. Investing.com México. Recuperado el 5 de diciembre de 2022, de <https://mx.investing.com/economic-calendar/cpi-733>

SPDR S&P 500 ETF Trust (SPY). (s/f). Yahoo.com. Recuperado el 5 de diciembre de 2022, de <https://finance.yahoo.com/quote/SPY/history?p=SPY>

Usd/mxn (mxn=x). (s/f). Yahoo.com. Recuperado el 5 de diciembre de 2022, de <https://finance.yahoo.com/quote/MXN%3DX/history?p=MXN%3DX>

Wikipedia contributors. (s/f). Método de Cochrane-Orcutt. Wikipedia, The Free Encyclopedia. https://es.wikipedia.org/w/index.php?title=M%C3%A9todo_de_Cochrane-Orcutt&oldid=117736472