

ANEXO 9

Regresión simple para todos los indicadores de las variables de entrada

María Camila Cardona^{a,c}, María José Martínez^{a,c}, Paula Andrea Bravo^{a,c}, Carlos Andrés Hernández^{a,c}

Isabel Cristina García^{b,c}

^aEstudiante de Ingeniería Industrial

^bProfesora, Directora del Proyecto de Grado, Departamento de Ciencias Naturales y Matemáticas

^cPontificia Universidad Javeriana, Cali, Colombia

Fecha de elaboración: 09/08/2021

La información presentada a través de este anexo, tiene la finalidad de conocer y analizar de forma más profunda el comportamiento de todas las variables de entrada de forma individual, por lo cual se evaluará la relación que tiene la tasa representativa del mercado (TRM), el depósito a término fijo (DTF), el precio del petróleo Brent, la tasa de desempleo en Estados Unidos y la tasa prime con respecto al índice COLCAP teniendo en cuenta un intervalo de tiempo desde el 25 de marzo del 2019 hasta el 25 de marzo del 2021

1. Evaluación de la tasa representativa del mercado (TRM)

Por medio de este modelo se pretende conocer cuál es el precio del índice COLCAP (indicador de la bolsa de valores de Colombia) con respecto a la tasa representativa del mercado (TRM). A partir de esto se realiza un análisis exploratorio por medio de diagramas de dispersión o nube de puntos para identificar puntos extraños.

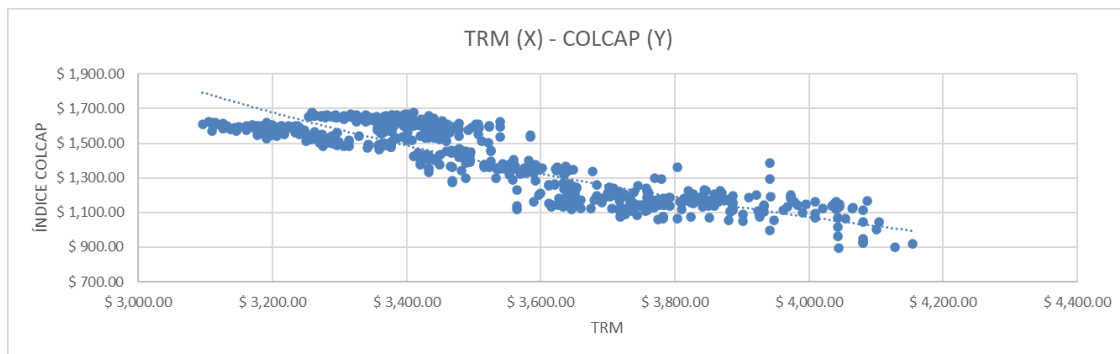


Fig. 107. Diagrama de dispersión del índice COLCAP y TRM

Se puede observar que existe una relación inversa, ya que a medida que aumenta el valor de la tasa representativa del mercado (TRM), el índice COLCAP disminuye. Por otra parte, no se observan puntos lejanos al patrón de comportamiento de la información que ha sido tomada.

a. Cálculo e interpretación del coeficiente de correlación (r).

El coeficiente de correlación indica el grado y el tipo de relación entre "X" y "Y". El grado de relación se refiere a: Si es alta, débil o no hay relación. Donde:

- $r \geq 0.8$, Relación alta
- $r < 0.8$, Relación débil

- $r = 0.8$, No hay relación

	TRM (X)	COLCAP (Y)
TRM (X)	1	
COLCAP (Y)	-0.887297691	1

Fig. 108. Coeficiente de correlación entre índice COLCAP y TRM

El coeficiente de correlación $r = -0.8873$, indica que hay una relación alta entre la TRM y el índice COLCAP y que hay una relación inversa entre el COLCAP y la TRM, ya que es negativo. A mayor TRM, menor tendrá que ser el valor del índice COLCAP.

- b. Identificación de la línea que mejor se ajusta a la serie de datos analizada.

Tipo de Modelo	R ²	Ecuación del modelo	fórmula general	Transformación	ANOVA
Lineal	0.7873	$y = -0.7569x + 4072.4$	$y = \beta_1 X + \beta_0$		$x ; y$
Exponencial	0.7791	$y = 10128e^{-6E-04x}$	$y = \beta_0 * e^{\beta_1 X}$	$\ln(y) = \ln(\beta_0) + \beta_1 X$	$x ; \ln(y)$
Polinómico ²	0.7877	$y = -6E-05 X^2 - 0.3117 X + 3279.3$	$y = \beta_2 X^2 + \beta_1 X + \beta_0$		$x ; x^2 ; y$
Potencial	0.7697	$y = 2E+10x^{-2.004}$	$y = \beta_0 X^{\beta_1}$	$\ln(y) = \ln(\beta_0) + \beta_1 \ln(x)$	$\ln(x) ; \ln(y)$
Logarítmico	0.7842	$y = -2694 \ln(x) + 23399$	$Y = \beta_1 \ln(x) + \beta_0$		$\ln(x) ; y$

Fig. 109. Comparación del R² con diferentes modelos

En este caso, se puede identificar que el coeficiente de determinación más alto lo tiene el modelo POLINÓMICO², por lo cual esta serie de datos tiene una relación curvilínea. Donde:

- X = Tasa representativa del mercado (TRM)
- Y = Índice COLCAP
- $\beta_0 = 3279.3$
- $\beta_1 = 0.3117$
- $\beta_2 = 6.00E-05$

Lo cual se resume a una fórmula más precisa $y = -6.00E - 05 X^2 - 0.3117 X + 3279.3$. El modelo de regresión planteado indica la variación de Y por unidad de cambio en X de forma polinómica.

- c. Estimación con el modelo de regresión \hat{Y} obtenido, reemplazando en el modelo un X de la base de datos.

$$y = \beta_2 X^2 + \beta_1 X + \beta_0$$

$$y = -6E-05 X^2 - 0.3117 X + 3279.3$$

Fig. 110. Fórmula para el modelo polinómico² con la TRM

Para realizar estimaciones con el modelo de regresión se debe tener en cuenta que no es válido salir del rango de datos de la variable X con la que se estimó el modelo de regresión.

Valor mínimo y máximo de la variable X	
Mínimo	\$ 3,095.66
Máximo	\$ 4,153.91

RANGO DE X	
\$ 3.095.66 < x < \$ 4,153.91	

Fig. 111. Rango de la tasa representativa del mercado (TRM)

Teniendo claro cuál es el rango de la variable X, se realiza el cálculo del \hat{Y} a partir de la formula planteada en la Fig. 110.

$\hat{Y} =$	\$ 1,466.43
$\beta_0 =$	3279.3
$\beta_1 =$	0.3117
$\beta_2 =$	6.00E-05
TRM (X) =	\$ 3,482.10

Fig. 112. Resultado del \hat{Y} con un valor de TRM (X)

El resultado anterior nos dice que si tengo un valor de TRM de \$3,482.1 pesos, se espera que el índice COLCAP de la bolsa de valores en Colombia este en \$1,466.43 pesos.

TRM (X)	COLCAP $Y_{\text{Observado}}$	$y = \beta_2 X^2 + \beta_1 X + \beta_0$ COLCAP $\hat{Y}_{\text{Estimado}}$	Residuo o error $Y_{\text{Observado}} - \hat{Y}_{\text{Estimado}}$
\$ 3,482.10	\$ 1,437.89	\$ 1,466.43	-\$ 28.54

Fig. 113. Comparación entre el Y observado y el Y estimado para la TRM

Al realizar la comparación entre los dos valores del índice COLCAP, se observa que el $\hat{Y}_{\text{Estimado}}$ teniendo en cuenta los coeficientes arrojados por la formula del modelo (β_0 , β_1 y β_2) es cercano al $Y_{\text{Observado}}$, en este caso con una diferencia de -\$28.54 pesos. A partir de esto, se podría decir que el modelo permite pronosticar el valor aproximado del índice COLCAP teniendo en cuenta el rango de valores para la variable TRM.

d. Realizar el análisis de varianza (ANOVA) teniendo en cuenta la transformación según el modelo escogido

Al realizar regresión en estos datos, se puede identificar nuevamente por medio de la Fig. 114 cuáles son los coeficientes de cada variable.

Resumen

Estadísticas de la regresión	
Coefficiente de correlación múltiple	0.887499359
Coefficiente de determinación R^2	0.787655111
R^2 ajustado	0.787072547
Error típico	94.29440175
Observaciones	732

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	2	24043275.42	12021637.71	1352.047087	5.0903E-246
Residuos	729	6481855.532	8891.434201		
Total	731	30525130.95			

	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 90.0%	Superior 90.0%
Intercepción	3279.267473	717.3649734	4.571267896	5.69369E-06	1870.919737	4687.61521	2097.80573	4460.72922
X	-0.311706795	0.401850434	-0.775678633	0.438190315	-1.100628988	0.4772154	-0.97353297	0.35011938
X^2	-6.21726E-05	5.60873E-05	-1.108497851	0.268012437	-0.000172284	4.7939E-05	-0.00015455	3.02E-05

Fig. 114. Regresión del modelo para la TRM

De igual forma en la tabla resumen, se puede observar cómo existe una relación alta entre la tasa representativa del mercado (TRM) y el índice COLCAP de la bolsa de valores de Colombia, ya que, el coeficiente de correlación es de 0.8874.

e. Interpretación del coeficiente de determinación R^2 obtenido.

Teniendo en cuenta el coeficiente de determinación R^2 el cual se puede visualizar en la Fig. 114, para este caso se puede decir que el 78.77% de la variación del índice COLCAP está siendo explicado por la tasa representativa del mercado (TRM) y el 21.23% restantes esta explicado por otras variaciones o factores que no han sido tomados en cuenta para este modelo.

Para mejora el porcentaje (%) de explicación del modelo (R^2), se deberían incluir otros factores que influyan en la variación de índice COLCAP, como el DTF, tasa prime, tasa de desempleo en USA, entre otras.

f. Interpretación de las sumas de cuadrados del ANOVA.

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Medio de los cuadrados	F	Valor crítico de F
Regresión	2	24,043,275.42	12021637.71	1352.047087	5.0903E-246
Residuos	729	6,481,855.53	8891.434201		
Total	731	30,525,130.95			

Fig. 115. Análisis de la suma de cuadrados para la TRM

Para saber si el modelo tiene un buen ajuste, se debe comparar la suma de cuadrados de error o residuo vs la suma de cuadrados de la regresión, si $SS_{Reg} > SS_{Residuos}$ indica que el modelo tiene un buen ajuste.

En este caso: $SS_{Reg} = 24,043,275.41 > SS_{Residuos} = 6,481,855.53$ lo que indica que el modelo tiene un buen ajuste. En la tabla de análisis de varianza, se puede observar que el modelo de variación total corresponde a 30,525,130.95. Donde se logra identificar que la variación explicada por el modelo es de 24,043,275.41 y el indicador de variación no explicada por el modelo o error es de 6,481,855.53.

g. Planteamiento y validación de la hipótesis correspondiente en la tabla ANOVA.

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Medio de los cuadrados	F	P-Valor
Regresión	2	24,043,275.42	12021637.71	1352.047087	5.09E-246
Residuos	729	6,481,855.53	8891.434201		
Total	731	30,525,130.95			

Fig. 116. Análisis del P-Valor para la TRM

Planteamiento de hipótesis para el modelo de regresión

Ho: $\beta_1 = 0$ vs **Ha:** $\beta_1 \neq 0$

Regla de decisión: Rechazo Ho si p-valor $< \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el p-valor = 5.09E-246 $< \alpha = 0.05$ **Rechazo Ho**

Conclusión: Con un nivel de confianza del 95%, existe suficiente evidencia para afirmar que β_1 y β_2 son significativamente diferente de cero, es decir, el modelo si explica en gran medida el comportamiento de Y. Esto quiere decir que la variación del índice COLCAP en la bolsa de valores se ve afectada significativamente por la tasa representativa del mercado (TRM) entre \$3.095.66 y \$ 4,153.91.

h. Intervalos de confianza y predicción de \hat{Y} utilizando los coeficientes arrojados por el ANOVA reemplazando en el modelo una de las X de la base de datos.

Al tener una base de datos tan grande en la Fig. 117 sólo se mostrarán los intervalos para los primeros 10 datos encontrados en la base de datos y así poder realizar la interpretación de forma clara.

N° DE DATOS	X	Y _{observado}	Pronosticos Y [^]	Intervalo de confianza del 95% (Corto plazo)		Intervalo de predicción del 95% (Largo plazo)	
	TRM	COLCAP	PFIT1	Limite inferior	Limite superior	Limite inferior	Limite superior
1	\$ 3,635.12	\$ 1,320.11	\$ 1,324.62	\$ 1,314.94	\$ 1,334.30	\$ 1,139.25	\$ 1,509.99
2	\$ 3,589.82	\$ 1,325.33	\$ 1,359.09	\$ 1,349.49	\$ 1,368.69	\$ 1,173.72	\$ 1,544.46
3	\$ 3,553.34	\$ 1,319.82	\$ 1,386.66	\$ 1,377.23	\$ 1,396.09	\$ 1,201.30	\$ 1,572.02
4	\$ 3,553.34	\$ 1,335.96	\$ 1,386.66	\$ 1,377.23	\$ 1,396.09	\$ 1,201.30	\$ 1,572.02
5	\$ 3,553.34	\$ 1,344.07	\$ 1,386.66	\$ 1,377.23	\$ 1,396.09	\$ 1,201.30	\$ 1,572.02
6	\$ 3,553.34	\$ 1,348.70	\$ 1,386.66	\$ 1,377.23	\$ 1,396.09	\$ 1,201.30	\$ 1,572.02
7	\$ 3,569.45	\$ 1,324.54	\$ 1,374.51	\$ 1,364.99	\$ 1,384.02	\$ 1,189.14	\$ 1,559.87
8	\$ 3,578.02	\$ 1,325.59	\$ 1,368.03	\$ 1,358.47	\$ 1,377.58	\$ 1,182.66	\$ 1,553.39
9	\$ 3,553.51	\$ 1,343.56	\$ 1,386.53	\$ 1,377.10	\$ 1,395.97	\$ 1,201.17	\$ 1,571.89
10	\$ 3,575.63	\$ 1,350.14	\$ 1,369.83	\$ 1,360.29	\$ 1,379.38	\$ 1,184.47	\$ 1,555.20

Fig. 117. Intervalos de confianza y de predicción al 95% para la TRM

A partir de esto se puede concluir que con un nivel de confianza del 95% se espera a **CORTO PLAZO** que el índice COLCAP de la bolsa de valores de Colombia esté entre \$1,377.23 y \$1,396.09 pesos cuando el valor de la tasa representativa del mercado (TRM) sea de \$3,553.34 pesos.

Por otra parte, con un nivel de confianza del 95% se espera a **LARGO PLAZO** que con el mismo valor de la tasa representativa del mercado (TRM), el índice COLCAP de la bolsa de valores de Colombia oscile entre \$1,201.30 y \$1,572.02 pesos.

X: Tasa Representativa del Mercado (TRM)	Y _{Observado} : Índice COLCAP	Y _{Estimado} : Índice COLCAP	
		Limite inferior	Limite Superior
\$ 3,553.34	\$ 1,335.96	\$ 1,377.23	\$ 1,396.09

Fig. 118. Comparación entre el Y observado y el intervalo del Y estimado a corto plazo para la TRM

Cuando la variable X= \$3,553.34. El Y_{Observado} no se encuentra dentro del intervalo de predicción de corto plazo, esto nos indica que la predicción a corto plazo no es asertiva y el intervalo tiene poca amplitud.

X: Tasa Representativa del Mercado (TRM)	Y _{Observado} : Índice COLCAP	Y _{Estimado} : Índice COLCAP	
		Limite inferior	Limite Superior
\$ 3,553.34	\$ 1,335.96	\$ 1,201.30	\$ 1,572.02

Fig. 119. Comparación entre el Y observado y el intervalo del Y estimado a largo plazo para la TRM

Cuando la variable X= \$3,553.34. El Y_{Observado} se encuentra dentro del intervalo de predicción de largo plazo, esto nos indica que la predicción es asertiva, ya que, el intervalo al 95% tiene una amplitud entre \$1,201.20 y \$1,572.02 pesos. Sin embargo, el intervalo es muy amplio, por lo cual el rango de error del Y_{Estimado} puede ser mayor.

VALIDACIÓN DE SUPUESTOS DE ANOVA

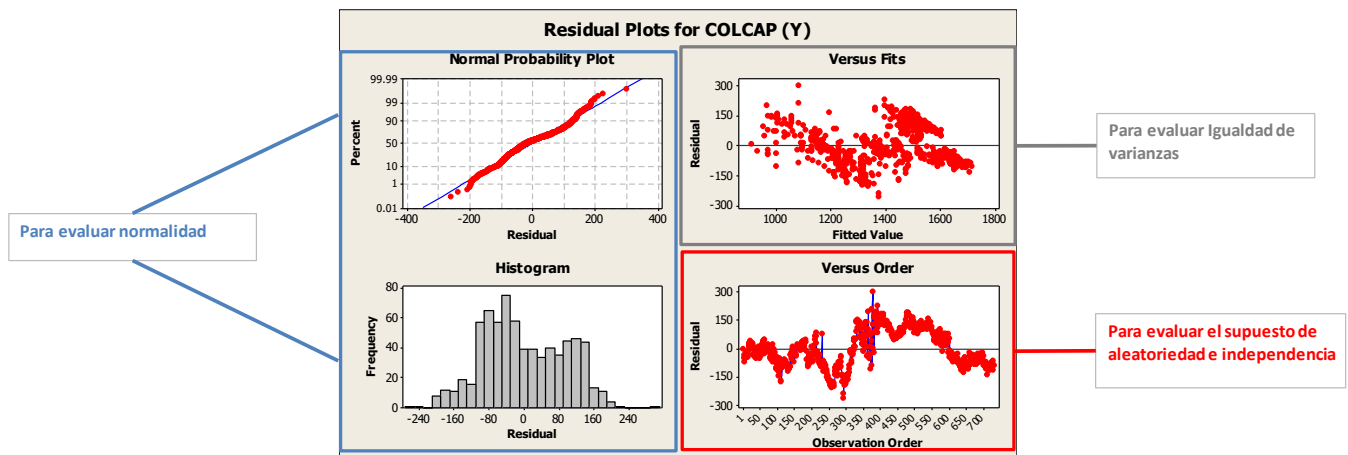


Fig. 120. Gráfico 4 en 1 para validación de supuestos de la TRM

El 4 en 1 nos muestra clara cuál es el comportamiento de los datos y ayuda a evaluar e identificar de forma gráfica los supuestos del ANOVA.

i. Planteamiento de hipótesis para verificar el cumplimiento del supuesto de normalidad de los residuos.

Planteamiento de hipótesis de prueba de bondad de ajuste a la distribución normal

H₀: Los residuos se ajustan a un comportamiento normal

H_a: Los residuos no se ajustan a un comportamiento normal

Regla de decisión: Rechazo H₀ si p-valor < α (para un nivel de confianza del 95%)

Decisión: Como el P-valor = 0.005 < $\alpha = 0.05$ **Rechazo H₀**

Conclusión: Con un nivel de confianza del 95%, se puede afirmar que los residuos NO se ajustan a un comportamiento normal. Lo que quiere decir, que hubo factores de ruido que afectaron la información tomada para el experimento y que dañaron el comportamiento normal en las variables.

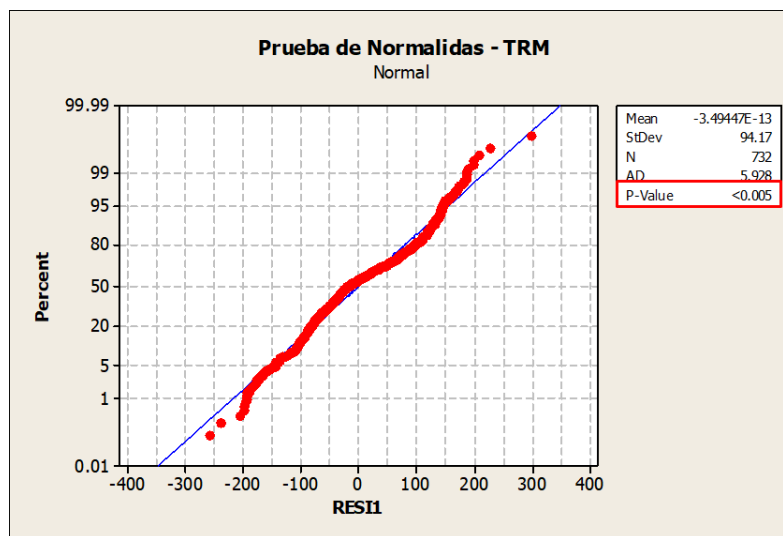


Fig. 121. Prueba de normalidad para los residuos de la TRM

j. Planteamiento de hipótesis para verificar el cumplimiento del supuesto de igualdad de varianzas.

Planteamiento de hipótesis de prueba de igualdad de varianzas "homocedasticidad"

H₀: Los residuos se tienen varianzas estadísticamente iguales

H_a: Los residuos no tienen varianzas estadísticamente iguales

Regla de decisión: Rechazo H₀ si p-valor < α (para un nivel de confianza del 95%)

Decisión: Como el P-valor = 0.027 < $\alpha = 0.05$ **Rechazo H₀**

Conclusión: Con un nivel de confianza del 95%, se puede afirmar que los residuos NO tienen varianzas estadísticamente iguales. Esto significa que las parejas de datos no oscilan aleatoriamente por encima y por debajo de cero con una variación estadísticamente igual, lo que permite decir que los datos no son comparables entre ellos y hay variables que tienen variación significativamente diferente.

Bartlett's Test (Normal Distribution)
 Test statistic = 433.76; p-value = 0.000

Levene's Test (Any Continuous Distribution)
 Test statistic = 1.35; p-value = 0.027

Fig. 122. P-Valor de la varianza obtenido en minitab para la TRM

- k. Planteamiento de hipótesis para verificar el cumplimiento del supuesto de aleatoriedad e independencia de los residuos.

Planteamiento de hipótesis de prueba de aleatoriedad e independencia

H₀: Los residuos son aleatorios e independientes

H_a: Los residuos no son aleatorios ni independientes

Regla de decisión:

Si $D.W. < dl$ Rechazo H_0

Si $dl < D.W. < du$ Cumple con reserva

Si $D.W. > du$ No rechazo H_0

Decisión: Como $D.W. = 0.130427 < dl = 1.748$ **Rechazo H_0**

Conclusión: Con un nivel de confianza del 95%, podemos afirmar que los residuos NO son aleatorios e independientes, lo que quiere decir que si existe autocorrelación y sesgos entre las variables analizadas.

Datos	
n =	732
k =	2
Nivel de confianza =	95%
dl =	1.748
du =	1.789

Fig. 123. Datos para encontrar el valor dl y du en tablas para la TRM

Dato arrojado en Minitab
Durbin-Watson statistic = 0.130427

Fig. 124. Prueba de aleatoriedad e independencia para la TRM

l. Conclusión

Después de plantear y validar las respectivas hipótesis para el supuesto de normalidad, igualdad de varianzas y aleatoriedad e independencia a un nivel de confianza del 95%, se puede decir que el modelo no es confiable para realizar estimaciones del índice COLCAP según la tasa representativa del mercado (TRM) a corto y largo plazo, ya que los 3 supuestos del ANOVA no se cumplen.

Por otra parte, con base al coeficiente de determinación (R^2 ajustado) el modelo mostró que el 78.77% de la variación del índice COLCAP está siendo explicado por él, y el 21.23% restante está explicado por otros factores que no han sido tomados en cuenta en el modelo.

2. Evaluación del depósito a término fijo (DTF)

Por medio de este modelo se pretende conocer cuál es el precio del índice COLCAP (indicador de la bolsa de valores de Colombia) con respecto a el depósito a término fijo (DTF). A partir de esto se realiza un análisis exploratorio por medio de diagramas de dispersión o nube de puntos para identificar puntos extraños.

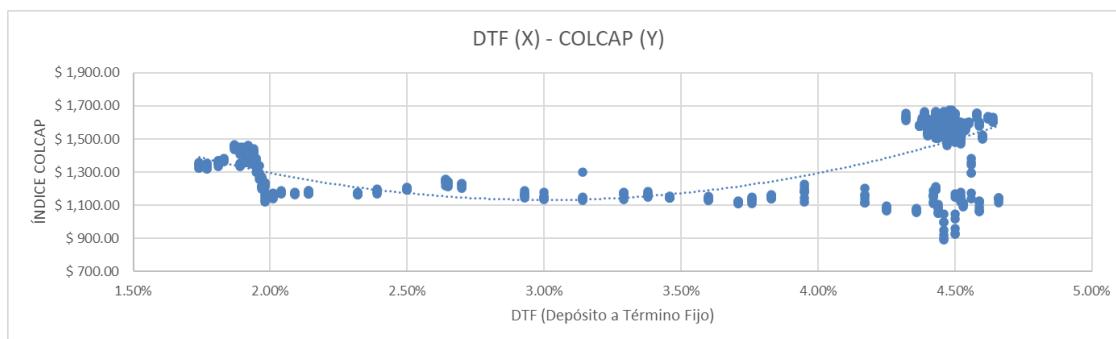


Fig. 125. Diagrama de dispersión del índice COLCAP y DTF

En el diagrama presentado no se observan puntos lejanos al patrón de comportamiento de la información que ha sido tomada.

a. Cálculo e interpretación el coeficiente de correlación (r).

El coeficiente de correlación indica el grado y el tipo de relación entre "X" y "Y". El grado de relación se refiere a: Si es alta, débil o no hay relación. Donde:

- $r \geq 0.8$, Relación alta
- $r < 0.8$, Relación débil
- $r = 0$, No hay relación

	DTF (X)	COLCAP (Y)
DTF (X)	1	
COLCAP (Y)	0.423943734	1

Fig. 126. Coeficiente de correlación entre índice COLCAP y DTF

El coeficiente de correlación $r = 0.4239$, indica que hay una relación baja entre el DTF y el índice COLCAP y que hay una relación directa entre el COLCAP y el DTF, ya que es positivo.

b. Identificación de la línea que mejor se ajusta a la serie de datos analizada.

Tipo de Modelo	R ²	Ecuación del modelo	fórmula general	Transformación	ANOVA
Lineal	0.1797	$y = 7895.1x + 1109.1$	$y = \beta_1 X + \beta_0$		$x ; y$
Exponencial	0.1845	$y = 1136e^{5.3667x}$	$y = \beta_0 * e^{\beta_1 X}$	$\ln(y) = \ln(\beta_0) + \beta_1 X$	$x ; \ln(y)$
Polinómico ²	0.3621	$y = 2E+06x^2 - 97654x + 2598.6$	$y = \beta_2 X^2 + \beta_1 X + \beta_0$		$x ; x^2 ; y$
Potencial	0.1549	$y = 2280.6x^{0.1488}$	$y = \beta_0 X^{\beta_1}$	$\ln(y) = \ln(\beta_0) + \beta_1 \ln(x)$	$\ln(x) ; \ln(y)$
Logarítmico	0.1505	$y = 220.92 \ln(x) + 2141.4$	$y = \beta_1 \ln(x) + \beta_0$		$\ln(x) ; y$

Fig. 127. Comparación del R² con diferentes modelos para el DTF

En este caso, se puede identificar que el coeficiente de determinación más alto lo tiene el modelo POLINÓMICO², por lo cual esta serie de datos tiene una relación curvilínea. Donde:

- X = Depósito a término fijo (DTF)
- Y = Índice COLCAP
- $\beta_0 = 2598.6$
- $\beta_1 = 97654$
- $\beta_2 = 2.00E+06$

Lo cual se resume a una fórmula más precisa $y = 2.00E + 06 x^2 - 97654x + 2598.6$. El modelo de regresión planteado indica la variación de Y por unidad de cambio en X de forma polinómica.

- c. Estimación con el modelo de regresión \hat{Y} obtenido, reemplazando en el modelo un X de la base de datos.

$$y = \beta_2 X^2 + \beta_1 X + \beta_0$$

$$y = 2E+06x^2 - 97654x + 2598.6$$

Fig. 128. Fórmula para el modelo polinómico² con el DTF

Para realizar estimaciones con el modelo de regresión se debe tener en cuenta que no es válido salir del rango de datos de la variable X con la que se estimó el modelo de regresión.

Valor mínimo y máximo de la variable X	
Mínimo	1.74%
Máximo	4.66%

RANGO DE X
$1.74\% < x < 4.66\%$

Fig. 129. Rango del depósito a término fijo (DTF)

Teniendo claro cuál es el rango de la variable X, se realiza el cálculo del \hat{Y} a partir de la formula planteada en la Fig. 128.

$\hat{Y} =$	\$ 1,467.36
$\beta_0 =$	2598.6
$\beta_1 =$	97654
$\beta_2 =$	2.00E+06
DTF (X) =	1.89%

Fig. 130. Resultado del \hat{Y} con un valor de DTF (X)

El resultado anterior nos dice que si tengo el valor del DTF de 1.89%, se espera que el índice COLCAP de la bolsa de valores en Colombia este en \$1,467.36 pesos.

		$y = \beta_2 X^2 + \beta_1 X + \beta_0$	Residuo o error
DTF (X)	COLCAP $Y_{\text{Observado}}$	COLCAP $\hat{Y}_{\text{Estimado}}$	$Y_{\text{Observado}} - \hat{Y}_{\text{Estimado}}$
1.89%	\$ 1,437.89	\$ 1,467.36	-\$ 29.47

Fig. 131. Comparación entre el Y observado y el Y estimado para el DTF

Al realizar la comparación entre los dos valores del índice COLCAP, se observa que el $\hat{Y}_{\text{Estimado}}$ teniendo en cuenta los coeficientes arrojados por la formula del modelo (β_0 , β_1 y β_2) es cercano al $Y_{\text{Observado}}$, en este caso con una diferencia de -\$29.47 pesos. A partir de esto, se podría decir que el modelo permite pronosticar el valor aproximado del índice COLCAP teniendo en cuenta el rango de valores para la variable DTF.

- d. Realizar el análisis de varianza (ANOVA) teniendo en cuenta la transformación según el modelo escogido

Al realizar regresión en estos datos, se puede identificar nuevamente por medio de la Fig. 132 cuáles son los coeficientes de cada variable.

Resumen

Estadísticas de la regresión	
Coefficiente de correlación múltiple	0.601771355
Coefficiente de determinación R ²	0.362128763
R ² ajustado	0.360378773
Error típico	163.4298994
Observaciones	732

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	2	11054027.92	5527013.958	206.9319426	6.68629E-72
Residuos	729	19471103.04	26709.33201		
Total	731	30525130.95			

	Coeficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 90.0%	Superior 90.0%
Intercepción	2598.591855	105.2823421	24.68212432	3.06909E-98	2391.899093	2805.28462	2425.19747	2771.98624
X	-97653.8063	7331.164179	-13.32036821	2.17584E-36	-112046.5197	-83261.0929	-109727.842	-85579.771
X2	1626399.914	112646.3458	14.4381063	9.60354E-42	1405249.967	1847549.86	1440877.41	1811922.42

Fig. 132. Regresión del modelo para la TRM

De igual forma en la tabla resumen, se puede observar cómo existe una relación baja entre el depósito a término fijo (DTF) y el índice COLCAP de la bolsa de valores de Colombia, ya que, el coeficiente de correlación es de 0.6018.

e. Interpretación del coeficiente de determinación R² obtenido.

Teniendo en cuenta el coeficiente de determinación R² el cual se puede visualizar en la Fig. 132, para este caso se puede decir que El 36.21% de la variación del índice COLCAP está siendo explicado por el depósito a término fijo (DTF). El 63.79% restantes esta explicado por otras variaciones o factores que no han sido tomados en cuenta para este modelo.

Para mejora el porcentaje (%) de explicación del modelo (R²), se deberían incluir otros factores que influyan en la variación de índice COLCAP, como la TRM, tasa prime, tasa de desempleo en USA, entre otras.

f. Interpretación de las sumas de cuadrados del ANOVA.

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	2	11,054,027.92	5527013.958	206.9319426	6.68629E-72
Residuos	729	19,471,103.04	26709.33201		
Total	731	30,525,130.95			

Fig. 133. Análisis de la suma de cuadrados para el DTF

Para saber si el modelo tiene un buen ajuste, se debe comparar la suma de cuadrados de error o residuo vs la suma de cuadrados de la regresión, si $SS_{Reg} > SS_{Residuos}$ indica que el modelo tiene un buen ajuste.

En este caso: $SS_{Reg} = 11,054,027.92 < SS_{Residuos} = 19,471,103.04$ lo que indica que el modelo NO tiene un buen ajuste. En la tabla de Análisis de varianza, se puede observar que el modelo de variación total corresponde a 30,525,130.95. Donde se logra identificar que la variación explicada por el modelo es de 11,054,027.92 y el indicador de variación no explicada por el modelo o error es de 19,471,103.04.

g. Planteamiento y validación de la hipótesis correspondiente en la tabla ANOVA.

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	P-Valor
Regresión	2	11,054,027.92	5527013.958	206.9319426	6.69E-72
Residuos	729	19,471,103.04	26709.33201		
Total	731	30,525,130.95			

Fig. 134. Análisis del P-Valor para el DTF

Planteamiento de hipótesis para el modelo de regresión

Ho: $\beta_1 = 0$ vs **Ha:** $\beta_1 \neq 0$

Regla de decisión: Rechazo Ho si $p\text{-valor} < \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el $p\text{-valor} = 6.69E-72 < \alpha = 0,05$ **Rechazo Ho**

Conclusión: Con un nivel de confianza del 95%, existe suficiente evidencia para afirmar que β_1 y β_2 son significativamente diferente de cero, es decir, el modelo sí explica en gran medida el comportamiento de Y. Esto quiere decir que la variación del índice COLCAP en la bolsa de valores se ve afectada significativamente por el depósito a término fijo (DTF) entre 1.74% y 4.66%

h. *Intervalos de confianza y predicción de Y^{\wedge} utilizando los coeficientes arrojados por el ANOVA reemplazando en el modelo una de las X de la base de datos.*

Al tener una base de datos tan grande en la Fig. 135 sólo se mostrarán los intervalos para los primeros 10 datos encontrados en la base de datos y así poder realizar la interpretación de forma clara.

N° DE DATOS	X	Y _{observado}	Pronosticos Y [^]	Intervalo de confianza del 95% (Corto plazo)		Intervalo de predicción del 95% (Largo plazo)	
	DTF	COLCAP	PFIT1	Limite inferior	Limite superior	Limite inferior	Limite superior
1	1.77%	\$ 1,320.11	\$ 1,379.65	\$ 1,350.11	\$ 1,409.20	\$ 1,057.45	\$ 1,701.86
2	1.77%	\$ 1,325.33	\$ 1,379.65	\$ 1,350.11	\$ 1,409.20	\$ 1,057.45	\$ 1,701.86
3	1.77%	\$ 1,319.82	\$ 1,379.65	\$ 1,350.11	\$ 1,409.20	\$ 1,057.45	\$ 1,701.86
4	1.77%	\$ 1,335.96	\$ 1,379.65	\$ 1,350.11	\$ 1,409.20	\$ 1,057.45	\$ 1,701.86
5	1.74%	\$ 1,344.07	\$ 1,391.82	\$ 1,360.84	\$ 1,422.81	\$ 1,069.48	\$ 1,714.17
6	1.74%	\$ 1,348.70	\$ 1,391.82	\$ 1,360.84	\$ 1,422.81	\$ 1,069.48	\$ 1,714.17
7	1.74%	\$ 1,324.54	\$ 1,391.82	\$ 1,360.84	\$ 1,422.81	\$ 1,069.48	\$ 1,714.17
8	1.74%	\$ 1,325.59	\$ 1,391.82	\$ 1,360.84	\$ 1,422.81	\$ 1,069.48	\$ 1,714.17
9	1.74%	\$ 1,343.56	\$ 1,391.82	\$ 1,360.84	\$ 1,422.81	\$ 1,069.48	\$ 1,714.17
10	1.74%	\$ 1,350.14	\$ 1,391.82	\$ 1,360.84	\$ 1,422.81	\$ 1,069.48	\$ 1,714.17

Fig. 135. Intervalos de confianza y de predicción al 95% para el DTF

A partir de esto se puede concluir que con un nivel de confianza del 95% se espera a **CORTO PLAZO** que el índice COLCAP de la bolsa de valores de Colombia esté entre \$1,350.11 y \$1,409.20 pesos cuando el valor del depósito a término fijo (DTF) sea de 1.77%.

Por otra parte, con un nivel de confianza del 95% se espera a **LARGO PLAZO** que con el mismo valor del DTF, el índice COLCAP de la bolsa de valores de Colombia oscile entre \$1,057.45 y \$1,701.86 pesos.

X: Depósito a término fijo (DTF)	Y _{Observado} : Índice COLCAP	Y _{Estimado} : Índice COLCAP	
		Limite inferior	Limite Superior
1.77%	\$ 1,335.96	\$ 1,350.11	\$ 1,409.20

Fig. 136. Comparación entre el Y observado y el intervalo del Y estimado a corto plazo para el DTF

Cuando la variable X= 1.77%. El Y_{Observado} no se encuentra dentro del intervalo de predicción de corto plazo, esto nos indica que la predicción a corto plazo no es asertiva y el intervalo tiene poca amplitud.

X: Depósito a término fijo (DTF)	Y _{Observado} : Índice COLCAP	Y _{Estimado} : Índice COLCAP	
		Limite inferior	Limite Superior
1.77%	\$ 1,335.96	\$ 1,057.45	\$ 1,701.86

Fig. 137. Comparación entre el Y observado y el intervalo del Y estimado a largo plazo para el DTF

Cuando la variable $X = 1.77\%$. El $Y_{\text{observado}}$ se encuentra dentro del intervalo de predicción de largo plazo, esto nos indica que la predicción es asertiva, ya que, el intervalo al 95% tiene una amplitud entre \$1,057.45 y \$1,701.86 Pesos. Sin embargo, el intervalo es muy amplio, por lo cual el rango de error del Y_{Estimado} puede ser mayor.

VALIDACIÓN DE SUPUESTOS DE ANOVA

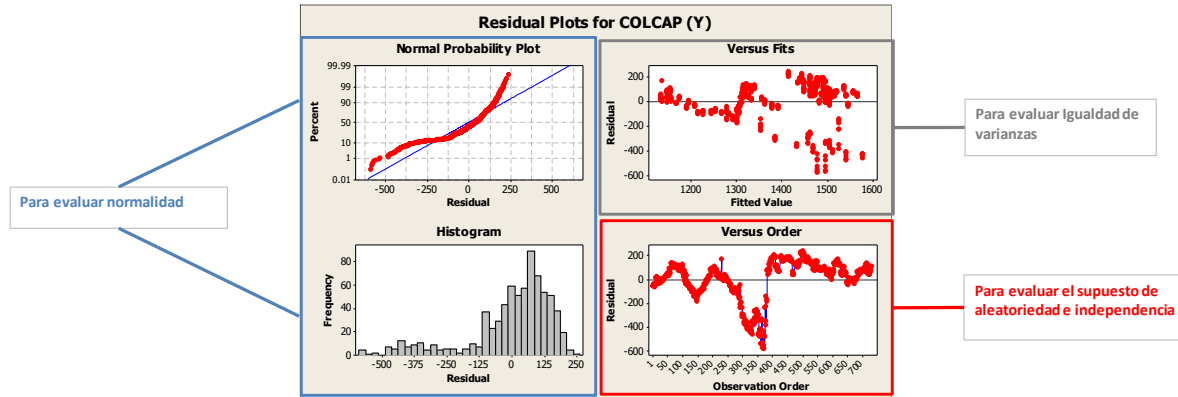


Fig. 138. Gráfico 4 en 1 para validación de supuestos del DTF

El 4 en 1 nos muestra clara cuál es el comportamiento de los datos y ayuda a evaluar e identificar de forma gráfica los supuestos del ANOVA.

- i. Planteamiento de hipótesis para verificar el cumplimiento del supuesto de normalidad de los residuos.

Planteamiento de hipótesis de prueba de bondad de ajuste a la distribución normal

H₀: Los residuos se ajustan a un comportamiento normal

H_a: Los residuos no se ajustan a un comportamiento normal

Regla de decisión: Rechazo H_0 si $p\text{-valor} < \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el $P\text{-valor} = 0.005 < \alpha = 0.05$ **Rechazo H_0**

Conclusión: Con un nivel de confianza del 95%, se puede afirmar que los residuos NO se ajustan a un comportamiento normal. Lo que quiere decir, que hubo factores de ruido que afectaron la información tomada para el experimento y que dañaron el comportamiento normal en las variables.

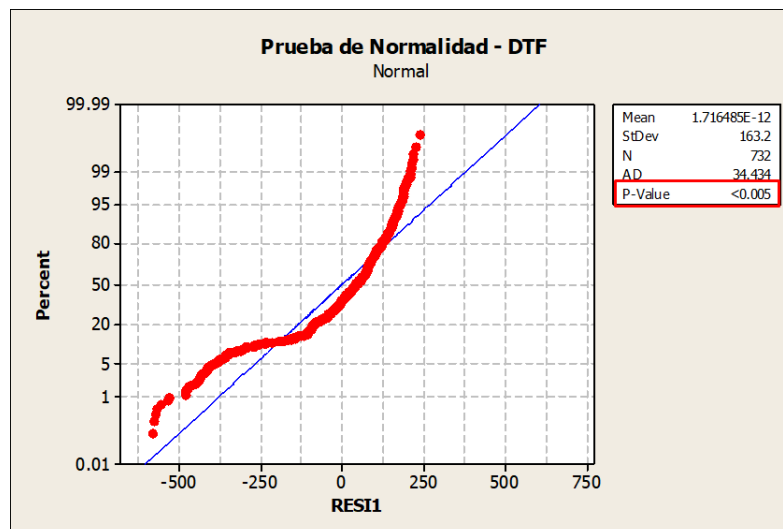


Fig. 139. Prueba de normalidad para los residuos del DTF

j. Planteamiento de hipótesis para verificar el cumplimiento del supuesto de igualdad de varianzas.

Planteamiento de hipótesis de prueba de igualdad de varianzas "homocedasticidad"

Ho: Los residuos se tienen varianzas estadísticamente iguales

Ha: Los residuos no tienen varianzas estadísticamente iguales

Regla de decisión: Rechazo Ho si $p\text{-valor} < \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el $P\text{-valor} = 0.000 < \alpha = 0.05$ **Rechazo Ho**

Conclusión: Con un nivel de confianza del 95%, se puede afirmar que los residuos NO tienen varianzas estadísticamente iguales. Esto significa que las parejas de datos no oscilan aleatoriamente por encima y por debajo de cero con una variación estadísticamente igual, lo que permite decir que los datos no son comparables entre ellos y hay variables que tienen variación significativamente diferente.

```

Bartlett's Test (Normal Distribution)
Test statistic = 1633.97; p-value = 0.000
  
```

```

Levene's Test (Any Continuous Distribution)
Test statistic = 5.34; p-value = 0.000
  
```

Fig. 140. P-Valor de la varianza obtenido en minitab para el DTF

k. Planteamiento de hipótesis para verificar el cumplimiento del supuesto de aleatoriedad e independencia de los residuos.

Planteamiento de hipótesis de prueba de aleatoriedad e independencia

Ho: Los residuos son aleatorios e independientes

Ha: Los residuos no son aleatorios ni independientes

Regla de decisión:

Si $D.W. < dl$ Rechazo Ho

Si $dl < D.W. < du$ Cumple con reserva

Si $D.W. > du$ No rechazo Ho

Decisión: Como $D.W. = 0.03927 < dl = 1.74$ **Rechazo Ho**

Conclusión: Con un nivel de confianza del 95%, podemos afirmar que los residuos NO son aleatorios e independientes, lo que quiere decir que si existe autocorrelación y sesgos entre las variables analizadas.

Datos	
n =	732
k =	2
Nivel de confianza =	95%
dl =	1.748
du =	1.789

Fig. 141. Datos para encontrar el valor dl y du en tablas para el DTF

Dato arrojado en Minitab
Durbin-Watson statistic = 0.0392696

Fig. 142. Prueba de aleatoriedad e independencia para el DTF

1. Conclusión

Después de plantear y validar las respectivas hipótesis para el supuesto de normalidad, igualdad de varianzas y aleatoriedad e independencia a un nivel de confianza del 95%, se puede decir que el modelo no es confiable para realizar estimaciones del índice COLCAP según el depósito a término fijo (DTF) a corto y largo plazo, ya que los 3 supuestos del ANOVA no se cumplen.

Por otra parte, con base al coeficiente de determinación (R^2 ajustado) el modelo mostró que el 36.21% de la variación del índice COLCAP está siendo explicado por él, y el 63.79% restante está explicado por otros factores que no han sido tomados en cuenta en el modelo.

3. Evaluación del precio del petróleo

Por medio de este modelo se pretende conocer cuál es el precio del índice COLCAP (indicador de la bolsa de valores de Colombia) con respecto a la tasa de desempleo en USA. A partir de esto se realiza un análisis exploratorio por medio de diagramas de dispersión o nube de puntos para identificar puntos extraños.

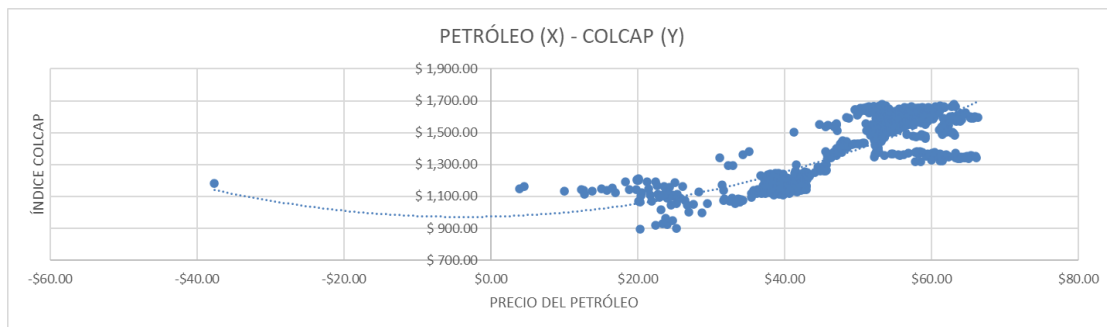


Fig. 143. Diagrama de dispersión del índice COLCAP y precio del petróleo

Se puede observar que tienen una relación directa. Por otra parte, se observa un punto lejano al patrón de comportamiento de la información que ha sido tomada por lo cual, se toma la decisión de eliminar ese dato y realizar nuevamente el diagrama.

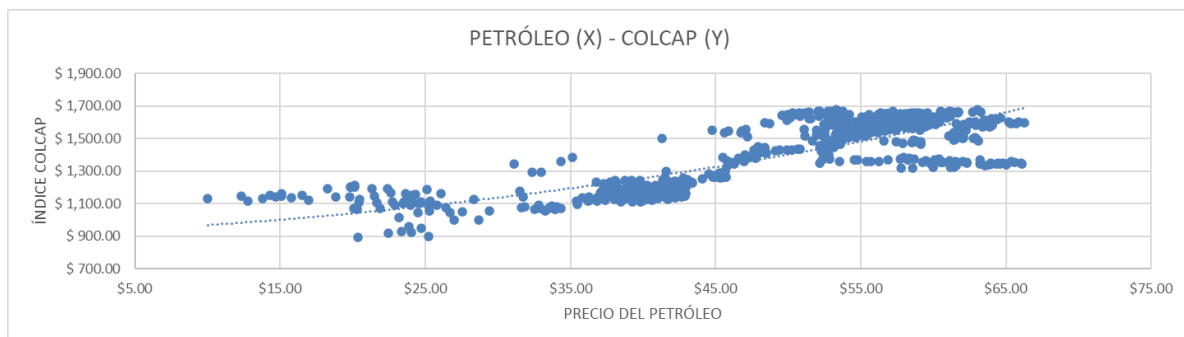


Fig. 144. Diagrama de dispersión del índice COLCAP y precio del petróleo corregido

a. Cálculo e interpretación el coeficiente de correlación (r).

El coeficiente de correlación indica el grado y el tipo de relación entre "X" y "Y". El grado de relación se refiere a: Si es alta, débil o no hay relación. Donde:

- $r \geq 0.8$, Relación alta
- $r < 0.8$, Relación débil
- $r = 0.8$, No hay relación

	PETRÓLEO (X)	COLCAP (Y)
PETRÓLEO (X)	1	
COLCAP (Y)	0.817538791	1

Fig. 145. Coeficiente de correlación entre índice COLCAP y el precio del petróleo

El coeficiente de correlación $r = 0.8175$, indica que hay una relación alta entre el precio del petróleo y el índice COLCAP y que hay una relación directa entre el COLCAP y el precio del petróleo, ya que es positivo.

b. Identificación de la línea que mejor se ajusta a la serie de datos analizada.

Tipo de Modelo	R ²	Ecuación del modelo	fórmula general	Transformación	ANOVA
Lineal	0.6684	$y = 14.046x + 713.54$	$y = \beta_1 X + \beta_0$		$x ; y$
Exponencial	0.6746	$y = 826.74e^{0.0106x}$	$y = \beta_0 * e^{\beta_1 X}$	$\ln(y) = \ln(\beta_0) + \beta_1 X$	$x ; \ln(y)$
Polinómico ²	0.6781	$y = 0.1196x^2 + 3.6293x + 920.3$	$y = \beta_2 X^2 + \beta_1 X + \beta_0$		$x ; x^2 ; y$
Potencial	0.6286	$y = 311.68x^{0.3872}$	$y = \beta_0 X^{\beta_1}$	$\ln(y) = \ln(\beta_0) + \beta_1 \ln(x)$	$\ln(x) ; \ln(y)$
Logaritmico	0.5864	$y = 513.78 \ln(x) - 578.51$	$Y = \beta_1 \ln(x) + \beta_0$		$\ln(x) ; y$

Fig. 146. Comparación del R² con diferentes modelos para el precio del petróleo

En este caso, se puede identificar que el coeficiente de determinación más alto lo tiene el modelo POLINÓMICO², por lo cual esta serie de datos tiene una relación curvilínea. Donde:

- X = Precio del petróleo
- Y = Índice COLCAP
- $\beta_0 = 920.3$
- $\beta_1 = 3.6293$
- $\beta_2 = 0.1196$

Lo cual se resume a una fórmula más precisa $y = 0.1196x^2 + 3.6293x + 920.3$. El modelo de regresión planteado indica la variación de Y por unidad de cambio en X de forma polinómica.

c. Estimación con el modelo de regresión \hat{Y} obtenido, reemplazando en el modelo un X de la base de datos.

$$y = \beta_2 X^2 + \beta_1 X + \beta_0$$

$$y = 0.1196x^2 + 3.6293x + 920.3$$

Fig. 147. Fórmula para el modelo polinómico² con el precio del petróleo

Para realizar estimaciones con el modelo de regresión se debe tener en cuenta que no es válido salir del rango de datos de la variable X con la que se estimó el modelo de regresión.

Valor mínimo y máximo de la variable X	
Mínimo	\$ 10.01
Máximo	\$ 66.30

RANGO DE X	
$\$10.01 < x < \66.30	

Fig. 148. Rango del precio del petróleo

Teniendo claro cuál es el rango de la variable X, se realiza el cálculo del \hat{Y} a partir de la formula planteada en la Fig. 147.

$\hat{Y} =$	\$ 1,376.13
$\beta_0 =$	920.3
$\beta_1 =$	3.6293
$\beta_2 =$	1.20E-01
Petróleo (X) =	\$ 48.40

Fig. 149. Resultado del \hat{Y} con un valor de precio de petróleo (X)

El resultado anterior nos dice que si tengo un precio de petróleo en \$48.40 dólares, se espera que el índice COLCAP de la bolsa de valores en Colombia este en \$1,376.13 pesos.

Petróleo (X)	COLCAP $Y_{\text{Observado}}$	$y = \beta_2 X^2 + \beta_1 X + \beta_0$	Residuo o error
		COLCAP $\hat{Y}_{\text{Estimado}}$	$Y_{\text{Observado}} - \hat{Y}_{\text{Estimado}}$
\$ 48.40	\$ 1,437.89	\$ 1,376.13	\$ 61.76

Fig. 150. Comparación entre el Y observado y el Y estimado para el precio del petróleo

Al realizar la comparación entre los dos valores del índice COLCAP, se observa que el $\hat{Y}_{\text{Estimado}}$ teniendo en cuenta los coeficientes arrojados por la formula del modelo (β_0 , β_1 y β_2) es cercano al $Y_{\text{Observado}}$, en este caso con una diferencia de \$61.76 pesos. A partir de esto, se podría decir que el modelo permite pronosticar el valor aproximado del índice COLCAP teniendo en cuenta el rango de valores para la variable de precio del petróleo.

d. Realizar el análisis de varianza (ANOVA) teniendo en cuenta la transformación según el modelo escogido

Al realizar regresión en estos datos, se puede identificar nuevamente por medio de la Fig. 151 cuáles son los coeficientes de cada variable.

Resumen

Estadísticas de la regresión	
Coefficiente de correlación múltiple	0.82272778
Coefficiente de determinación R^2	0.676881
R^2 ajustado	0.67599331
Error típico	116.3140261
Observaciones	731

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	2	20632192.62	10316096.31	762.5199502	2.5583E-179
Residuos	728	9849077.54	13528.95266		
Total	730	30481270.16			

	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 90.0%	Superior 90.0%
Intercepción	927.6941971	47.8266901	19.39699768	6.97578E-68	833.7995032	1021.58889	848.926059	1006.46234
X	3.254552232	2.258199061	1.441215829	0.149953697	-1.178807239	7.6879117	-0.46458735	6.97369182
X2	0.123654655	0.025586433	4.832821185	1.64243E-06	0.073422655	0.17388665	0.08151509	0.16579421

Fig. 151. Regresión del modelo para el precio del petróleo

De igual forma en la tabla resumen, se puede observar cómo existe una relación alta entre el precio del petróleo y el índice COLCAP de la bolsa de valores de Colombia, ya que, el coeficiente de correlación es de 0.8227

e. Interpretación del coeficiente de determinación R^2 obtenido.

Teniendo en cuenta el coeficiente de determinación R^2 el cual se puede visualizar en la Fig. 151, para este caso se puede decir que el 67.69% de la variación del índice COLCAP está siendo explicado por el precio del petróleo. El 32.31% restantes esta explicado por otras variaciones o factores que no han sido tomados en cuenta para este modelo.

Para mejorar el porcentaje (%) de explicación del modelo (R^2), se deberían incluir otros factores que influyan en la variación de índice COLCAP, como el DTF, tasa prime, tasa de desempleo en USA, entre otras.

f. Interpretación de las sumas de cuadrados del ANOVA.

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Medio de los cuadrados	F	Valor crítico de F
Regresión	2	20,632,192.62	10316096.31	762.5199502	2.5583E-179
Residuos	728	9,849,077.54	13528.95266		
Total	730	30,481,270.16			

Fig. 152. Análisis de la suma de cuadrados para el precio del petróleo

Para saber si el modelo tiene un buen ajuste, se debe comparar la suma de cuadrados de error o residuo vs la suma de cuadrados de la regresión, si $SS_{Reg} > SS_{Residuos}$ indica que el modelo tiene un buen ajuste.

En este caso: $SS_{Reg} = 20,632,192.64 > SS_{Residuos} = 9,849,077.54$ lo que indica que el modelo tiene un buen ajuste. En la tabla de Análisis de varianza, se puede observar que el modelo de variación total corresponde a 30,481,270.16. Donde se logra identificar que la variación explicada por el modelo es de 20,632,192.64 y el indicador de variación no explicada por el modelo o error es de 9,849,077.54.

g. Planteamiento y validación de la hipótesis correspondiente en la tabla ANOVA.

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Medio de los cuadrados	F	P-Valor
Regresión	2	20,632,192.62	10316096.31	762.5199502	2.56E-179
Residuos	728	9,849,077.54	13528.95266		
Total	730	30,481,270.16			

Fig. 153. Análisis del P-Valor para el precio del petróleo

Planteamiento de hipótesis para el modelo de regresión

Ho: $\beta_1 = 0$ vs **Ha:** $\beta_1 \neq 0$

Regla de decisión: Rechazo H_0 si $p\text{-valor} < \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el $p\text{-valor} = 2.56E-179 < \alpha = 0.05$ **Rechazo H_0**

Conclusión: Con un nivel de confianza del 95%, existe suficiente evidencia para afirmar que β_1 y β_2 son significativamente diferente de cero, es decir, el modelo si explica en gran medida el comportamiento de Y. Esto quiere decir que la variación del índice COLCAP en la bolsa de valores se ve afectada significativamente por el precio del petróleo entre \$10.01 y \$66.30 dólares.

h. Intervalos de confianza y predicción de Y^A utilizando los coeficientes arrojados por el ANOVA reemplazando en el modelo una de las X de la base de datos.

Al tener una base de datos tan grande en la Fig. 154 sólo se mostrarán los intervalos para los primeros 10 datos encontrados en la base de datos y así poder realizar la interpretación de forma clara.

N° DE DATOS				Intervalo de confianza del 95% (Corto plazo)		Intervalo de predicción del 95% (Largo plazo)	
	X	Y observado	Pronósticos Y^A	Límite inferior	Límite superior	Límite inferior	Límite superior
	Petróleo	COLCAP	PFIT1	CLIM1	CLIM2	PLIM1	PLIM2
1	\$58.56	\$ 1,320.11	\$ 1,542.33	\$ 1,531.00	\$ 1,553.65	\$ 1,313.69	\$ 1,770.96
2	\$61.18	\$ 1,325.33	\$ 1,589.65	\$ 1,575.48	\$ 1,603.82	\$ 1,360.86	\$ 1,818.44
3	\$57.76	\$ 1,319.82	\$ 1,528.22	\$ 1,517.48	\$ 1,538.95	\$ 1,299.61	\$ 1,756.82
4	\$61.55	\$ 1,335.96	\$ 1,596.47	\$ 1,581.79	\$ 1,611.14	\$ 1,367.64	\$ 1,825.29
5	\$63.24	\$ 1,344.07	\$ 1,628.08	\$ 1,610.79	\$ 1,645.38	\$ 1,399.08	\$ 1,857.09
6	\$63.96	\$ 1,348.70	\$ 1,641.77	\$ 1,623.22	\$ 1,660.32	\$ 1,412.67	\$ 1,870.88
7	\$61.42	\$ 1,324.54	\$ 1,594.07	\$ 1,579.57	\$ 1,608.56	\$ 1,365.26	\$ 1,822.88
8	\$60.00	\$ 1,325.59	\$ 1,568.12	\$ 1,555.40	\$ 1,580.84	\$ 1,339.42	\$ 1,796.83
9	\$64.60	\$ 1,343.56	\$ 1,653.97	\$ 1,634.25	\$ 1,673.69	\$ 1,424.77	\$ 1,883.17
10	\$64.80	\$ 1,350.14	\$ 1,657.82	\$ 1,637.72	\$ 1,677.92	\$ 1,428.59	\$ 1,887.05

Fig. 154. Intervalos de confianza y de predicción al 95% para el precio del petróleo

A partir de esto se puede concluir que con un nivel de confianza del 95% se espera a **CORTO PLAZO** que el índice COLCAP de la bolsa de valores de Colombia esté entre \$1,581.79 y \$1,611.14 pesos cuando el valor del precio del petróleo sea de \$61.55.

Por otra parte, con un nivel de confianza del 95% se espera a **LARGO PLAZO** que, con el mismo valor del precio del petróleo, el índice COLCAP de la bolsa de valores de Colombia oscile entre \$1,367.64 y \$1,825.29 pesos.

X: Precio del petróleo	Y _{Observado} : Índice COLCAP	Y _{Estimado} : índice COLCAP	
		Limite inferior	Limite Superior
\$ 61.55	\$ 1,335.96	\$ 1,581.79	\$ 1,611.14

Fig. 155. Comparación entre el Y observado y el intervalo del Y estimado a corto plazo para el precio del petróleo

Cuando la variable X= \$61.55. El Y_{Observado} no se encuentra dentro del intervalo de predicción de corto plazo, esto nos indica que la predicción a corto plazo no es asertiva y el intervalo tiene poca amplitud.

X1: Precio del petróleo	Y _{Observado} : Índice COLCAP	Y _{Estimado} : Índice COLCAP	
		Limite inferior	Limite Superior
\$ 61.55	\$ 1,335.96	\$ 1,367.64	\$ 1,825.29

Fig. 156. Comparación entre el Y observado y el intervalo del Y estimado a largo plazo para el precio del petróleo

Cuando la variable X= \$61.55. El Y_{Observado} no se encuentra dentro del intervalo de predicción de largo plazo, esto nos indica que la predicción no es asertiva, ya que, el intervalo al 95% tiene una amplitud entre \$1,367.64 y \$1,825.29 Pesos.

VALIDACIÓN DE SUPUESTOS DE ANOVA

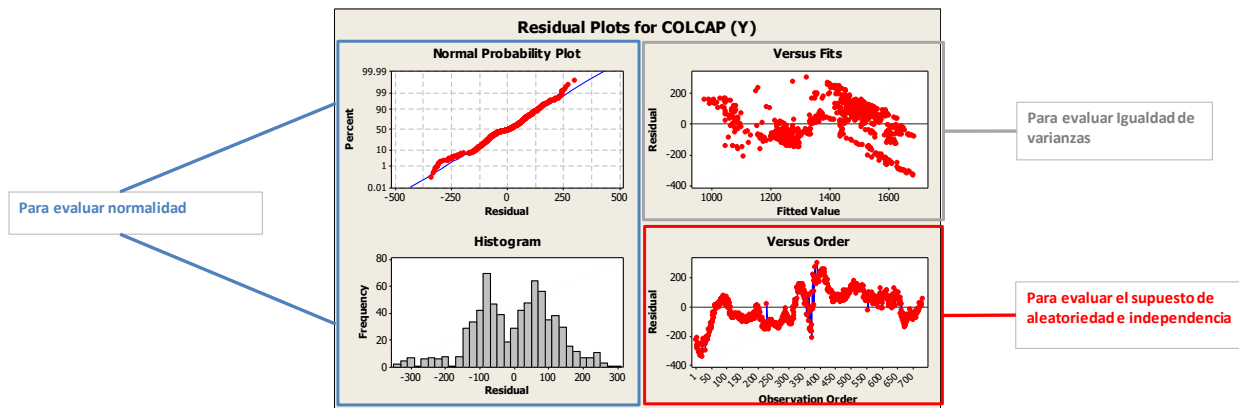


Fig. 157. Gráfico 4 en 1 para validación de supuestos del precio del petróleo

El 4 en 1 nos muestra clara cuál es el comportamiento de los datos y ayuda a evaluar e identificar de forma gráfica los supuestos del ANOVA.

- Planteamiento de hipótesis para verificar el cumplimiento del supuesto de normalidad de los residuos.

Planteamiento de hipótesis de prueba de bondad de ajuste a la distribución normal

H₀: Los residuos se ajustan a un comportamiento normal

Ha: Los residuos no se ajustan a un comportamiento normal

Regla de decisión: Rechazo H_0 si $p\text{-valor} < \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el $P\text{-valor} = 0.005 < \alpha = 0.05$ **Rechazo H_0**

Conclusión: Con un nivel de confianza del 95%, se puede afirmar que los residuos NO se ajustan a un comportamiento normal. Lo que quiere decir, que hubo factores de ruido que afectaron la información tomada para el experimento y que dañaron el comportamiento normal en las variables.

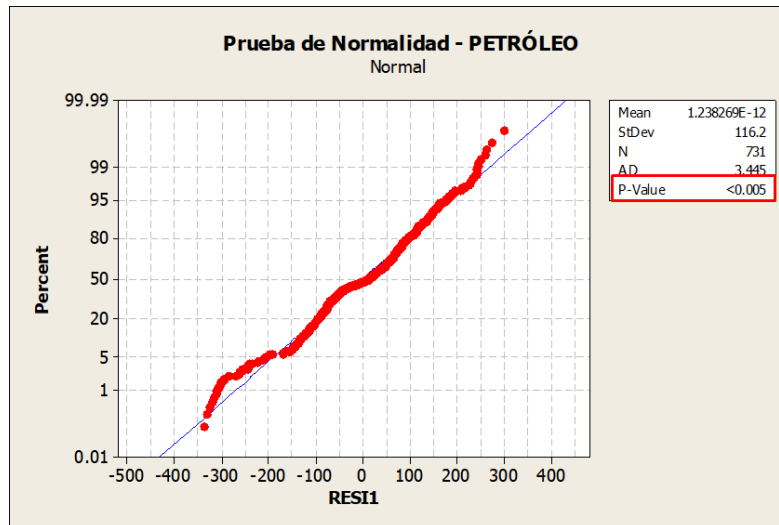


Fig. 158. Prueba de normalidad para los residuos del precio del petróleo

j. Planteamiento de hipótesis para verificar el cumplimiento del supuesto de igualdad de varianzas.

Planteamiento de hipótesis de prueba de igualdad de varianzas "homocedasticidad"

H_0 : Los residuos se tienen varianzas estadísticamente iguales

Ha: Los residuos no tienen varianzas estadísticamente iguales

Regla de decisión: Rechazo H_0 si $p\text{-valor} < \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el $P\text{-valor} = 0.000 < \alpha = 0.05$ **Rechazo H_0**

Conclusión: Con un nivel de confianza del 95%, se puede afirmar que los residuos NO tienen varianzas estadísticamente iguales. Esto significa que las parejas de datos no oscilan aleatoriamente por encima y por debajo de cero con una variación estadísticamente igual, lo que permite decir que los datos no son comparables entre ellos y hay variables que tienen variación significativamente diferente.

Bartlett's Test (Normal Distribution)
Test statistic = 97.39; p-value = 0.000

Levene's Test (Any Continuous Distribution)
Test statistic = 4.24; p-value = 0.000

Fig. 159. P-Valor de la varianza obtenido en minitab para el precio del petróleo

k. Planteamiento de hipótesis para verificar el cumplimiento del supuesto de aleatoriedad e independencia de los residuos.

Planteamiento de hipótesis de prueba de aleatoriedad e independencia

H₀: Los residuos son aleatorios e independientes

H_a: Los residuos no son aleatorios ni independientes

Regla de decisión:

Si $D.W. < d_l$ Rechazo H_0

Si $d_l < D.W. < d_u$ Cumple con reserva

Si $D.W. > d_u$ No rechazo H_0

Decisión: Como $D.W. = 0.0962 < d_l = 1.748$ **Rechazo H_0**

Conclusión: Con un nivel de confianza del 95%, podemos afirmar que los residuos NO son aleatorios e independientes, lo que quiere decir que si existe autocorrelación y sesgos entre las variables analizadas.

Datos	
n =	731
k =	2
Nivel de confianza =	95%
d_l =	1.748
d_u =	1.789

Fig. 160. Datos para encontrar el valor d_l y d_u en tablas para el precio del petróleo

Dato arrojado en Minitab
Durbin-Watson statistic = 0.0962299

Fig. 161. Prueba de aleatoriedad e independencia para el precio del petróleo

1. Conclusión

Después de plantear y validar las respectivas hipótesis para el supuesto de normalidad, igualdad de varianzas y aleatoriedad e independencia a un nivel de confianza del 95%, se puede decir que el modelo no es confiable para realizar estimaciones del índice COLCAP según el precio del petróleo a corto y largo plazo, ya que los 3 supuestos del ANOVA no se cumplen.

Por otra parte, con base al coeficiente de determinación (R^2 ajustado) el modelo mostró que el 67.69 % de la variación del índice COLCAP está siendo explicado por él, y el 32.31% restante está explicado por otros factores que no han sido tomados en cuenta en el modelo.

4. Evaluación de la tasa de desempleo en USA

Por medio de este modelo se pretende conocer cuál es el precio del índice COLCAP (indicador de la bolsa de valores de Colombia) con respecto a la tasa de desempleo en USA. A partir de esto se realiza un análisis exploratorio por medio de diagramas de dispersión o nube de puntos para identificar puntos extraños.

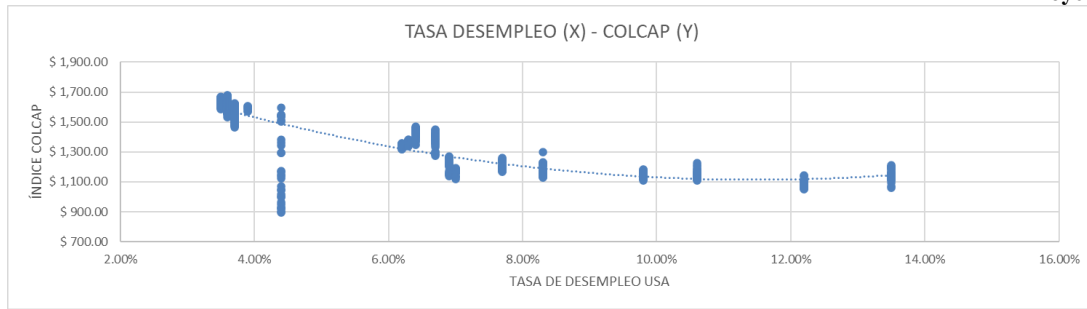


Fig. 162. Diagrama de dispersión del índice COLCAP y la tasa de desempleo

En el diagrama presentado no se observan puntos lejanos al patrón de comportamiento de la información que ha sido tomada.

a. *Cálculo e interpretación el coeficiente de correlación (r).*

El coeficiente de correlación indica el grado y el tipo de relación entre "X" y "Y". El grado de relación se refiere a: Si es alta, débil o no hay relación. Donde:

- $r \geq 0.8$, Relación alta
- $r < 0.8$, Relación débil
- $r = 0$, No hay relación

	TASA DESEMPLEO (X)	COLCAP (Y)
TASA DESEMPLEO (X)	1	
COLCAP (Y)	-0.828680569	1

Fig. 163. Coeficiente de correlación entre índice COLCAP y la tasa de desempleo en USA

El coeficiente de correlación $r = 0.8175$, indica que hay una relación alta entre la tasa de desempleo en USA y el índice COLCAP y que hay una relación inversa entre el COLCAP y la tasa de desempleo en USA, ya que es negativo.

b. *Identificación de la línea que mejor se ajusta a la serie de datos analizada.*

Tipo de Modelo	R ²	Ecuación del modelo	fórmula general	Transformación	ANOVA
Lineal	0.6867	$y = -5774.4x + 1747.5$	$y = \beta_1 X + \beta_0$		$x ; y$
Exponencial	0.7162	$y = 1787.9e^{-4.249x}$	$y = \beta_0 * e^{\beta_1 X}$	$\ln(y) = \ln(\beta_0) + \beta_1 X$	$x ; \ln(y)$
Polinómico ²	0.7849	$y = 74652x^2 - 17134x + 2097.5$	$y = \beta_2 X^2 + \beta_1 X + \beta_0$		$x ; x^2 ; y$
Potencial	0.7772	$y = 584.5x^{-0.296}$	$y = \beta_0 X^{\beta_1}$	$\ln(y) = \ln(\beta_0) + \beta_1 \ln(x)$	$\ln(x) ; \ln(y)$
Logaritmico	0.7644	$y = -404.7 \ln(x) + 221.35$	$y = \beta_1 \ln(x) + \beta_0$		$\ln(x) ; y$

Fig. 164. Comparación del R² con diferentes modelos para la tasa de desempleo

En este caso, se puede identificar que el coeficiente de determinación más alto lo tiene el modelo POLINÓMICO², por lo cual esta serie de datos tiene una relación curvilínea. Donde:

- X = Tasa de desempleo en USA
- Y = Índice COLCAP
- $\beta_0 = 2097.5$
- $\beta_1 = 17134$
- $\beta_2 = 74652$

Lo cual se resume a una fórmula más precisa $y = 74652x^2 - 17134x + 2097.5$. El modelo de regresión planteado indica la variación de Y por unidad de cambio en X de forma polinómica.

c. *Estimación con el modelo de regresión Y^ obtenido, reemplazando en el modelo un X de la base de datos.*

$$y = \beta_2 X^2 + \beta_1 X + \beta_0$$

$$y = 74652x^2 - 17134x + 2097.5$$

Fig. 165. Fórmula para el modelo polinómico² con la tasa de desempleo en USA

Para realizar estimaciones con el modelo de regresión se debe tener en cuenta que no es válido salir del rango de datos de la variable X con la que se estimó el modelo de regresión.

Valor mínimo y máximo de la variable X	
Mínimo	3.50%
Máximo	13.50%

RANGO DE X
3.50% < x < 13.50%

Fig. 166. Rango de la tasa de desempleo en USA

Teniendo claro cuál es el rango de la variable X, se realiza el cálculo del \hat{Y} a partir de la formula planteada en la Fig. 165.

$\hat{Y} =$	\$ 1,284.63
$\beta_0 =$	2097.5
$\beta_1 =$	17134
$\beta_2 =$	7.47E+04
Tasa de Desempleo (X) =	6.70%

Fig. 167. Resultado del \hat{Y} con un valor de la tasa de desempleo en USA (X)

El resultado anterior nos dice que si la tasa de desempleo es de 6.70%, se espera que el índice COLCAP de la bolsa de valores en Colombia este en \$1,284.63.

Tasa de Desempleo	COLCAP $Y_{\text{Observado}}$	$y = \beta_2 X^2 + \beta_1 X + \beta_0$	Residuo o error
		COLCAP $\hat{Y}_{\text{Estimado}}$	$Y_{\text{Observado}} - \hat{Y}_{\text{Estimado}}$
6.70%	\$ 1,437.89	\$ 1,284.63	\$ 153.26

Fig. 168. Comparación entre el Y observado y el Y estimado para la tasa de desempleo en USA

Al realizar la comparación entre los dos valores del índice COLCAP, se observa que el $\hat{Y}_{\text{Estimado}}$ teniendo en cuenta los coeficientes arrojados por la formula del modelo (β_0 , β_1 y β_2) es cercano al $Y_{\text{Observado}}$, en este caso con una diferencia de \$153.26 pesos. A partir de esto, se podría decir que el modelo permite pronosticar el valor aproximado del índice COLCAP teniendo en cuenta el rango de valores para la variable de la tasa de desempleo en USA.

d. Realizar el análisis de varianza (ANOVA) teniendo en cuenta la transformación según el modelo escogido

Al realizar regresión en estos datos, se puede identificar nuevamente por medio de la Fig. 169 cuáles son los coeficientes de cada variable.

Resumen

Estadísticas de la regresión	
Coefficiente de correlación múltiple	0.885922327
Coefficiente de determinación R ²	0.78485837
R ² ajustado	0.784268132
Error típico	94.91333451
Observaciones	732

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	2	23957904.51	11978952.26	1329.732769	5.9988E-244
Residuos	729	6567226.439	9008.541068		
Total	731	30525130.95			

	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 90.0%	Superior 90.0%
Intercepción	2097.529438	20.81573316	100.7665415	0	2056.663503	2138.39537	2063.24704	2131.81184
X	-17133.56672	634.2826426	-27.01251078	6.6293E-112	-18378.80528	-15888.3282	-18178.1963	-16088.9371
X2	74652.01145	4093.565281	18.23642872	1.68839E-61	66615.42811	82688.5948	67910.1284	81393.8945

Fig. 169. Regresión del modelo para la tasa de desempleo en USA

De igual forma en la tabla resumen, se puede observar cómo existe una relación alta entre la tasa de desempleo en USA y el índice COLCAP de la bolsa de valores de Colombia, ya que, el coeficiente de correlación es de 0.8859

e. Interpretación del coeficiente de determinación R² obtenido.

Teniendo en cuenta el coeficiente de determinación R² el cual se puede visualizar en la Fig. 169, para este caso se puede decir que el 78.49% de la variación del índice COLCAP está siendo explicado por la tasa de desempleo en USA. El 21.51% restantes esta explicado por otras variaciones o factores que no han sido tomados en cuenta para este modelo.

Para mejora el porcentaje (%) de explicación del modelo (R²), se deberían incluir otros factores que influyan en la variación de índice COLCAP, como el DTF, tasa prime, precio del petróleo, entre otras.

f. Interpretación de las sumas de cuadrados del ANOVA.

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	2	23,957,904.51	11978952.26	1329.732769	5.9988E-244
Residuos	729	6,567,226.44	9008.541068		
Total	731	30,525,130.95			

Fig. 170. Análisis de la suma de cuadrados para la tasa de desempleo en USA

Para saber si el modelo tiene un buen ajuste, se debe comparar la suma de cuadrados de error o residuo vs la suma de cuadrados de la regresión, si $SS_{Reg} > SS_{Residuos}$ indica que el modelo tiene un buen ajuste.

En este caso: $SS_{Reg} = 23,957,904.51 > SS_{Residuos} = 6,567,226.44$ lo que indica que el modelo tiene un buen ajuste. En la tabla de Análisis de varianza, se puede observar que el modelo de variación total corresponde a 30,525,130.95. Donde se logra identificar que la variación explicada por el modelo es de 23,957,904.51 y el indicador de variación no explicada por el modelo o error es de 6,567,226.44.

g. Planteamiento y validación de la hipótesis correspondiente en la tabla ANOVA.

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	P-Valor
Regresión	2	23,957,904.51	11978952.26	1329.732769	6.00E-244
Residuos	729	6,567,226.44	9008.541068		
Total	731	30,525,130.95			

Fig. 171. Análisis del P-Valor para la tasa de desempleo en USA

Planteamiento de hipótesis para el modelo de regresión

H₀: $\beta_1 = 0$ vs **H_a:** $\beta_1 \neq 0$

Regla de decisión: Rechazo H_0 si $p\text{-valor} < \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el $p\text{-valor} = 6.00E-244 < \alpha = 0,05$ **Rechazo H_0**

Conclusión: Con un nivel de confianza del 95%, existe suficiente evidencia para afirmar que β_1 y β_2 son significativamente diferente de cero, es decir, el modelo si explica en gran medida el comportamiento de Y. Esto quiere decir que la variación del índice COLCAP en la bolsa de valores se ve afectada significativamente por la tasa de desempleo en USA entre 3.50% y 13.50%

h. *Intervalos de confianza y predicción de Y^{\wedge} utilizando los coeficientes arrojados por el ANOVA reemplazando en el modelo una de las X de la base de datos.*

Al tener una base de datos tan grande en la Fig. 172 sólo se mostrarán los intervalos para los primeros 10 datos encontrados en la base de datos y así poder realizar la interpretación de forma clara.

N° DE DATOS	X	Y _{observado} COLCAP	Pronosticos Y [^] PFIT1	Intervalo de confianza del 95% (Corto plazo)		Intervalo de predicción del 95% (Largo plazo)	
				Limite inferior	Limite superior	Limite inferior	Limite superior
	Desempleo			CLIM1	CLIM2	PLIM1	PLIM2
1	6.20%	\$ 1,320.11	\$ 1,322.21	\$ 1,312.21	\$ 1,332.21	\$ 1,135.61	\$ 1,508.81
2	6.20%	\$ 1,325.33	\$ 1,322.21	\$ 1,312.21	\$ 1,332.21	\$ 1,135.61	\$ 1,508.81
3	6.20%	\$ 1,319.82	\$ 1,322.21	\$ 1,312.21	\$ 1,332.21	\$ 1,135.61	\$ 1,508.81
4	6.20%	\$ 1,335.96	\$ 1,322.21	\$ 1,312.21	\$ 1,332.21	\$ 1,135.61	\$ 1,508.81
5	6.20%	\$ 1,344.07	\$ 1,322.21	\$ 1,312.21	\$ 1,332.21	\$ 1,135.61	\$ 1,508.81
6	6.20%	\$ 1,348.70	\$ 1,322.21	\$ 1,312.21	\$ 1,332.21	\$ 1,135.61	\$ 1,508.81
7	6.20%	\$ 1,324.54	\$ 1,322.21	\$ 1,312.21	\$ 1,332.21	\$ 1,135.61	\$ 1,508.81
8	6.20%	\$ 1,325.59	\$ 1,322.21	\$ 1,312.21	\$ 1,332.21	\$ 1,135.61	\$ 1,508.81
9	6.20%	\$ 1,343.56	\$ 1,322.21	\$ 1,312.21	\$ 1,332.21	\$ 1,135.61	\$ 1,508.81
10	6.20%	\$ 1,350.14	\$ 1,322.21	\$ 1,312.21	\$ 1,332.21	\$ 1,135.61	\$ 1,508.81

Fig. 172. Intervalos de confianza y de predicción al 95% para la tasa de desempleo en USA

A partir de esto se puede concluir que con un nivel de confianza del 95% se espera a **CORTO PLAZO** que el índice COLCAP de la bolsa de valores de Colombia esté entre \$1,312.21 y \$1,332.21 pesos cuando el valor de la tasa de desempleo en USA sea de 6.20%.

Por otra parte, con un nivel de confianza del 95% se espera a **LARGO PLAZO** que, con el mismo valor de la tasa de desempleo en USA, el índice COLCAP de la bolsa de valores de Colombia oscile entre \$1,135.61 y \$1,508.81 pesos.

X: Tasa de desempleo en USA	Y _{Observado} : Índice COLCAP	Y _{Estimado} : índice COLCAP	
		Limite inferior	Limite Superior
6.20%	\$ 1,335.96	\$ 1,312.21	\$ 1,332.21

Fig. 173. Comparación entre el Y observado y el intervalo del Y estimado a corto plazo para la tasa de desempleo

Cuando la variable $X = 6.20\%$. El $Y_{\text{Observado}}$ no se encuentra dentro del intervalo de predicción de corto plazo, esto nos indica que la predicción a corto plazo no es asertiva y el intervalo tiene poca amplitud.

X: Tasa de desempleo en USA	Y _{Observado} : Índice COLCAP	Y _{Estimado} : Índice COLCAP	
		Limite inferior	Limite Superior
6.20%	\$ 1,335.96	\$ 1,135.61	\$ 1,508.81

Fig. 174. Comparación entre el Y observado y el intervalo del Y estimado a largo plazo para la tasa de desempleo

Cuando la variable $X = 6.20\%$. El $Y_{\text{Observado}}$ se encuentra dentro del intervalo de predicción de largo plazo, esto nos indica que la predicción es asertiva, ya que, el intervalo al 95% tiene una amplitud entre \$1,135.61 y \$1,508.81 Pesos. Sin embargo, el intervalo es muy amplio, por lo cual el rango de error del Y_{Estimado} puede ser mayor.

VALIDACIÓN DE SUPUESTOS DE ANOVA

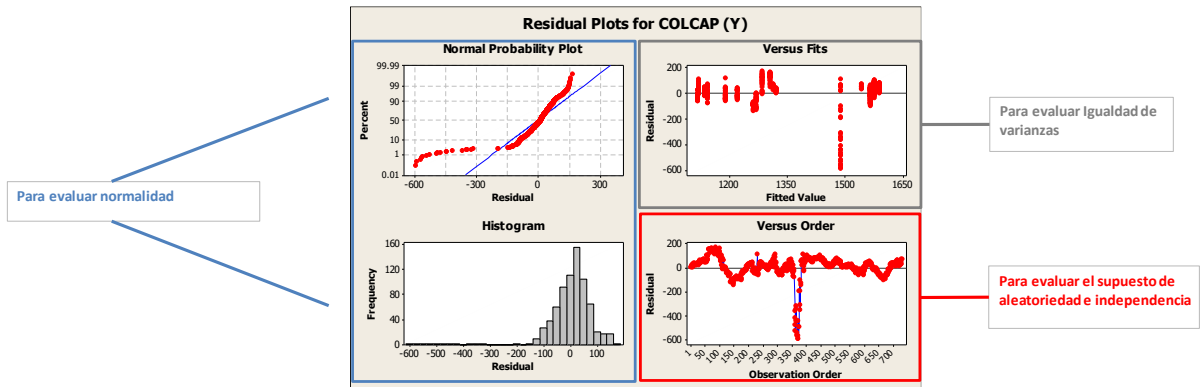


Fig. 175. Gráfico 4 en 1 para validación de supuestos de la tasa de desempleo en USA

El 4 en 1 nos muestra clara cuál es el comportamiento de los datos y ayuda a evaluar e identificar de forma gráfica los supuestos del ANOVA.

i. Planteamiento de hipótesis para verificar el cumplimiento del supuesto de normalidad de los residuos.

Planteamiento de hipótesis de prueba de bondad de ajuste a la distribución normal

Ho: Los residuos se ajustan a un comportamiento normal

Ha: Los residuos no se ajustan a un comportamiento normal

Regla de decisión: Rechazo Ho si $p\text{-valor} < \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el $P\text{-valor} = 0.005 < \alpha = 0.05$ **Rechazo Ho**

Conclusión: Con un nivel de confianza del 95%, se puede afirmar que los residuos NO se ajustan a un comportamiento normal. Lo que quiere decir, que hubo factores de ruido que afectaron la información tomada para el experimento y que dañaron el comportamiento normal en las variables.

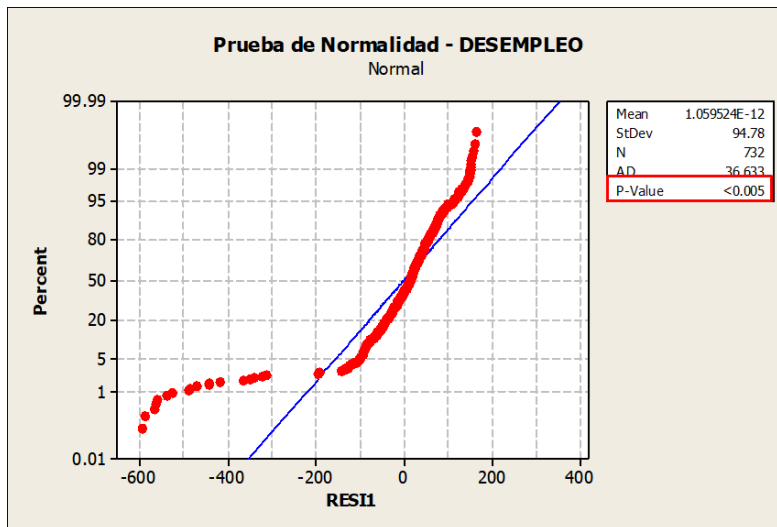


Fig. 176. Prueba de normalidad para los residuos de la tasa de desempleo en USA

j. *Planteamiento de hipótesis para verificar el cumplimiento del supuesto de igualdad de varianzas.*

Planteamiento de hipótesis de prueba de igualdad de varianzas "homocedasticidad"

Ho: Los residuos se tienen varianzas estadísticamente iguales

Ha: Los residuos no tienen varianzas estadísticamente iguales

Regla de decisión: Rechazo Ho si $p\text{-valor} < \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el $P\text{-valor} = 0.000 < \alpha = 0.05$ **Rechazo Ho**

Conclusión: Con un nivel de confianza del 95%, se puede afirmar que los residuos NO tienen varianzas estadísticamente iguales. Esto significa que las parejas de datos no oscilan aleatoriamente por encima y por debajo de cero con una variación estadísticamente igual, lo que permite decir que los datos no son comparables entre ellos y hay variables que tienen variación significativamente diferente.

```
Bartlett's Test (Normal Distribution)
Test statistic = 857.89; p-value = 0.000
```

```
Levene's Test (Any Continuous Distribution)
Test statistic = 59.13; p-value = 0.000
```

Fig. 177. P-Valor de la varianza obtenido en minitab para la tasa de desempleo en USA

k. *Planteamiento de hipótesis para verificar el cumplimiento del supuesto de aleatoriedad e independencia de los residuos.*

Planteamiento de hipótesis de prueba de aleatoriedad e independencia

Ho: Los residuos son aleatorios e independientes

Ha: Los residuos no son aleatorios ni independientes

Regla de decisión:

Si $D.W. < dl$ Rechazo Ho

Si $dl < D.W. < du$ Cumple con reserva

Si $D.W. > du$ No rechazo Ho

Decisión: Como $D.W. = 0.105517 < dl = 1.748$ **Rechazo Ho**

Conclusión: Con un nivel de confianza del 95%, podemos afirmar que los residuos NO son aleatorios e independientes, lo que quiere decir que si existe autocorrelación y sesgos entre las variables analizadas.

Datos	
n =	732
k =	2
Nivel de confianza =	95%
dl =	1.748
du =	1.789

Fig. 178. Datos para encontrar el valor dl y du en tablas para la tasa de desempleo en USA

Dato arrojado en Minitab
Durbin-Watson statistic = 0.105517

Fig. 179. Prueba de aleatoriedad e independencia para la tasa de desempleo en USA

1. Conclusión

Después de plantear y validar las respectivas hipótesis para el supuesto de normalidad, igualdad de varianzas y aleatoriedad e independencia a un nivel de confianza del 95%, se puede decir que el modelo no es confiable para realizar estimaciones del índice COLCAP según la tasa de desempleo en USA a corto y largo plazo, ya que los 3 supuestos del ANOVA no se cumplen.

Por otra parte, con base al coeficiente de determinación (R^2 ajustado) el modelo mostró que el 78.49% de la variación del índice COLCAP está siendo explicado por él, y el 21.51% restante está explicado por otros factores que no han sido tomados en cuenta en el modelo.

5. Evaluación de la tasa prime

Por medio de este modelo se pretende conocer cuál es el precio del índice COLCAP (indicador de la bolsa de valores de Colombia) con respecto a la tasa prime. A partir de esto se realiza un análisis exploratorio por medio de diagramas de dispersión o nube de puntos para identificar puntos extraños.

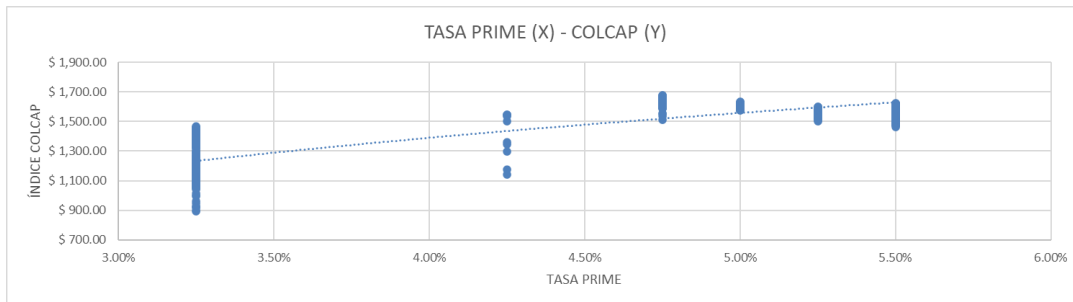


Fig. 180. Diagrama de dispersión del índice COLCAP y la tasa prime

En el diagrama presentado no se observan puntos lejanos al patrón de comportamiento de la información que ha sido tomada.

a. Cálculo e interpretación el coeficiente de correlación (r).

El coeficiente de correlación indica el grado y el tipo de relación entre "X" y "Y". El grado de relación se refiere a: Si es alta, débil o no hay relación. Donde:

- $r \geq 0.8$, Relación alta
- $r < 0.8$, Relación débil
- $r = 0.8$, No hay relación

	TASA PRIME (X)	COLCAP (Y)
TASA PRIME (X)	1	
COLCAP (Y)	0.834993044	1

Fig. 181. Coeficiente de correlación entre índice COLCAP y la tasa prime

El coeficiente de correlación $r = 0.83499$, indica que hay una relación alta entre la tasa prime y el índice COLCAP y que hay una relación directa entre el COLCAP y la tasa prime, ya que es positivo.

b. Identificación de la línea que mejor se ajusta a la serie de datos analizada.

Tipo de Modelo	R ²	Ecuación del modelo	fórmula general	Transformación	ANOVA
Lineal	0.6972	$y = 17754x + 660.85$	$y = \beta_1 X + \beta_0$		x ; y
Exponencial	0.6822	$y = 806.92e^{12.967x}$	$y = \beta_0 * e^{\beta_1 X}$	$\ln(y) = \ln(\beta_0) + \beta_1 X$	x ; ln(y)
Polinómico ²	0.7887	$y = -1E+06x^2 + 145453x - 1924.6$	$y = \beta_2 X^2 + \beta_1 X + \beta_0$		x ; x ² ; y
Potencial	0.7077	$y = 8062x^{0.5496}$	$y = \beta_0 X^{\beta_1}$	$\ln(y) = \ln(\beta_0) + \beta_1 \ln(x)$	ln(x) ; ln(y)
Logaritmico	0.7191	$y = 753.12 \ln(x) + 3814.1$	$Y = \beta_1 \ln(x) + \beta_0$		ln(x) ; y

Fig. 182. Comparación del R² con diferentes modelos para la tasa prime

En este caso, se puede identificar que el coeficiente de determinación más alto lo tiene el modelo POLINÓMICO², por lo cual esta serie de datos tiene una relación curvilínea. Donde:

- X = Tasa Prime
- Y = Índice COLCAP
- $\beta_0 = 1924.6$
- $\beta_1 = 145453$
- $\beta_2 = 1.00E+06$

Lo cual se resume a una fórmula más precisa $y = -1E + 06x^2 + 145453x - 1924.6$. El modelo de regresión planteado indica la variación de Y por unidad de cambio en X de forma polinómica.

c. Estimación con el modelo de regresión Y[^] obtenido, reemplazando en el modelo un X de la base de datos.

$$y = \beta_2 X^2 + \beta_1 X + \beta_0$$

$$y = -1E+06x^2 + 145453x - 1924.6$$

Fig. 183. Fórmula para el modelo polinómico² con la tasa prime

Para realizar estimaciones con el modelo de regresión se debe tener en cuenta que no es válido salir del rango de datos de la variable X con la que se estimó el modelo de regresión.

Valor mínimo y máximo de la variable X	
Mínimo	3.25%
Máximo	5.50%

RANGO DE X
$3.25\% < x < 5.50\%$

Fig. 184. Rango de la tasa prime

Teniendo claro cuál es el rango de la variable X, se realiza el cálculo del Y[^] a partir de la formula planteada en la Fig. 183.

Y [^] =	\$ 1,746.37
β_0 =	1924.6
β_1 =	145453
β_2 =	1.00E+06
Tasa Prime (X) =	3.25%

Fig. 185. Resultado del Y[^] con un valor de la tasa prime (X)

El resultado anterior nos dice que si la tasa prime es de 3.25%, se espera que el índice COLCAP de la bolsa de valores en Colombia este en \$1,746.37.

TASA PRIME (X)	COLCAP Y _{Observado}	$y = \beta_2 X^2 + \beta_1 X + \beta_0$	Residuo o error
		COLCAP Y [^] _{Estimado}	Y _{Observado} - Y [^] _{Estimado}
3.25%	\$ 1,437.89	\$ 1,746.37	-\$ 308.48

Fig. 186. Comparación entre el Y observado y el Y estimado para la tasa prime

Al realizar la comparación entre los dos valores del índice COLCAP, se observa que el Y_{Estimado} teniendo en cuenta los coeficientes arrojados por la fórmula del modelo (β_0 , β_1 y β_2) es cercano al $Y_{\text{Observado}}$, en este caso con una diferencia de -\$308.48 pesos. A partir de esto, se podría decir que el modelo permite pronosticar el valor aproximado del índice COLCAP teniendo en cuenta el rango de valores para la variable de la tasa prime.

d. Realizar el análisis de varianza (ANOVA) teniendo en cuenta la transformación según el modelo escogido

Al realizar regresión en estos datos, se puede identificar nuevamente por medio de la Fig. 187 cuáles son los coeficientes de cada variable.

Resumen

Estadísticas de la regresión	
Coefficiente de correlación múltiple	0.888112984
Coefficiente de determinación R^2	0.788744673
R^2 ajustado	0.788165098
Error típico	94.05217384
Observaciones	732

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	2	24076534.44	12038267.22	1360.90028	7.8054E-247
Residuos	729	6448596.514	8845.811404		
Total	731	30525130.95			

	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%	Inferior 90.0%	Superior 90.0%
Intercepción	-1924.573902	146.2894543	-13.15593055	1.27082E-35	-2211.772788	-1637.37501	-2165.50481	-1683.64299
X	145452.7546	7194.368852	20.21758372	1.68532E-72	131328.601	159576.908	133604.014	157301.495
X2	-1495792.62	84164.05738	-17.77234447	5.58376E-59	-1661025.471	-1330559.77	-1634406.32	-1357178.92

Fig. 187. Regresión del modelo para la tasa de desempleo en USA

De igual forma en la tabla resumen, se puede observar cómo existe una relación alta entre la tasa prime y el índice COLCAP de la bolsa de valores de Colombia, ya que, el coeficiente de correlación es de 0.8881

e. Interpretación del coeficiente de determinación R^2 obtenido.

Teniendo en cuenta el coeficiente de determinación R^2 el cual se puede visualizar en la Fig. 187, para este caso se puede decir que El 78.87% de la variación del índice COLCAP está siendo explicado por la tasa prime. El 21.13% restantes esta explicado por otras variaciones o factores que no han sido tomados en cuenta para este modelo.

Para mejora el porcentaje (%) de explicación del modelo (R^2), se deberían incluir otros factores que influyan en la variación de índice COLCAP, como el DTF, tasa de desempleo en USA, precio del petróleo, entre otras.

f. Interpretación de las sumas de cuadrados del ANOVA.

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Promedio de los cuadrados	F	Valor crítico de F
Regresión	2	24,076,534.44	12038267.22	1360.90028	7.8054E-247
Residuos	729	6,448,596.51	8845.811404		
Total	731	30,525,130.95			

Fig. 188. Análisis de la suma de cuadrados para la tasa prime

Para saber si el modelo tiene un buen ajuste, se debe comparar la suma de cuadrados de error o residuo vs la suma de cuadrados de la regresión, si $SS_{\text{Reg}} > SS_{\text{Residuos}}$ indica que el modelo tiene un buen ajuste.

En este caso: $SS_{\text{Reg}} = 24,076,534.44 > SS_{\text{Residuos}} = 6,448,596.51$ lo que indica que el modelo tiene un buen ajuste. En la tabla de Análisis de varianza, se puede observar que el modelo de variación total corresponde a 30,525,130.95. Donde se logra identificar que la variación explicada por el modelo es de 24,076,534.44 y el indicador de variación no explicada por el modelo o error es de 6,448,596.51

g. Planteamiento y validación de la hipótesis correspondiente en la tabla ANOVA.

ANÁLISIS DE VARIANZA

	Grados de libertad	Suma de cuadrados	Medio de los cuadrados	F	P-Valor
Regresión	2	24,076,534.44	12038267.22	1360.90028	7.81E-247
Residuos	729	6,448,596.51	8845.811404		
Total	731	30,525,130.95			

Fig. 189. Análisis del P-Valor para la tasa prime

Planteamiento de hipótesis para el modelo de regresión

Ho: $\beta_1 = 0$ vs **Ha:** $\beta_1 \neq 0$

Regla de decisión: Rechazo Ho si p-valor $< \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el p-valor = $7.81E-247 < \alpha = 0.05$ **Rechazo Ho**

Conclusión: Con un nivel de confianza del 95%, existe suficiente evidencia para afirmar que β_1 y β_2 son significativamente diferente de cero, es decir, el modelo sí explica en gran medida el comportamiento de Y. Esto quiere decir que la variación del índice COLCAP en la bolsa de valores se ve afectada significativamente por la tasa prime entre 3.25% y 5.50%.

h. Intervalos de confianza y predicción de \hat{Y} utilizando los coeficientes arrojados por el ANOVA reemplazando en el modelo una de las X de la base de datos.

Al tener una base de datos tan grande en la Fig. 190 sólo se mostrarán los intervalos para los primeros 10 datos encontrados en la base de datos y así poder realizar la interpretación de forma clara.

N° DE DATOS	X	Y _{observado}	Pronosticos Y ^A	Intervalo de confianza del 95% (Corto plazo)		Intervalo de predicción del 95% (Largo plazo)	
	PRIME	COLCAP	PFIT1	Limite inferior	Limite superior	Limite inferior	Limite superior
1	3.25%	\$ 1,320.11	\$ 1,222.71	\$ 1,213.21	\$ 1,232.21	\$ 1,037.82	\$ 1,407.60
2	3.25%	\$ 1,325.33	\$ 1,222.71	\$ 1,213.21	\$ 1,232.21	\$ 1,037.82	\$ 1,407.60
3	3.25%	\$ 1,319.82	\$ 1,222.71	\$ 1,213.21	\$ 1,232.21	\$ 1,037.82	\$ 1,407.60
4	3.25%	\$ 1,335.96	\$ 1,222.71	\$ 1,213.21	\$ 1,232.21	\$ 1,037.82	\$ 1,407.60
5	3.25%	\$ 1,344.07	\$ 1,222.71	\$ 1,213.21	\$ 1,232.21	\$ 1,037.82	\$ 1,407.60
6	3.25%	\$ 1,348.70	\$ 1,222.71	\$ 1,213.21	\$ 1,232.21	\$ 1,037.82	\$ 1,407.60
7	3.25%	\$ 1,324.54	\$ 1,222.71	\$ 1,213.21	\$ 1,232.21	\$ 1,037.82	\$ 1,407.60
8	3.25%	\$ 1,325.59	\$ 1,222.71	\$ 1,213.21	\$ 1,232.21	\$ 1,037.82	\$ 1,407.60
9	3.25%	\$ 1,343.56	\$ 1,222.71	\$ 1,213.21	\$ 1,232.21	\$ 1,037.82	\$ 1,407.60
10	3.25%	\$ 1,350.14	\$ 1,222.71	\$ 1,213.21	\$ 1,232.21	\$ 1,037.82	\$ 1,407.60

Fig. 190. Intervalos de confianza y de predicción al 95% para la tasa prime

A partir de esto se puede concluir que con un nivel de confianza del 95% se espera a **CORTO PLAZO** que el índice COLCAP de la bolsa de valores de Colombia esté entre \$1,213.21 y \$1,232.21 pesos cuando el valor de la tasa prime sea de 3.25%.

Por otra parte, con un nivel de confianza del 95% se espera a **LARGO PLAZO** que, con el mismo valor de la tasa prime, el índice COLCAP de la bolsa de valores de Colombia oscile entre \$1,037.82 y \$1,407.60 pesos.

X: Tasa Prime	Y _{Observado} : Índice COLCAP	Y _{Estimado} : índice COLCAP	
		Limite inferior	Limite Superior
3.25%	\$ 1,335.96	\$ 1,213.21	\$ 1,232.21

Fig. 191. Comparación entre el Y observado y el intervalo del Y estimado a corto plazo para la tasa prime

Cuando la variable X= 3.25%. El Y_{Observado} no se encuentra dentro del intervalo de predicción de corto plazo, esto nos indica que la predicción a corto plazo no es asertiva y el intervalo tiene poca amplitud.

X: Tasa Prime	Y _{Observado} : Índice COLCAP	Y _{Estimado} : Índice COLCAP	
		Limite inferior	Limite Superior
3.25%	\$ 1,335.96	\$ 1,037.82	\$ 1,407.60

Fig. 192. Comparación entre el Y observado y el intervalo del Y estimado a largo plazo para la tasa prime

Cuando la variable $X = 3.25\%$. El $Y_{\text{Observado}}$ se encuentra dentro del intervalo de predicción de largo plazo, esto nos indica que la predicción es asertiva, ya que, el intervalo al 95% tiene una amplitud entre \$1,037.82 y \$1,407.60 pesos. Sin embargo, el intervalo es muy amplio, por lo cual el rango de error del Y_{Estimado} puede ser mayor.

VALIDACIÓN DE SUPUESTOS DE ANOVA

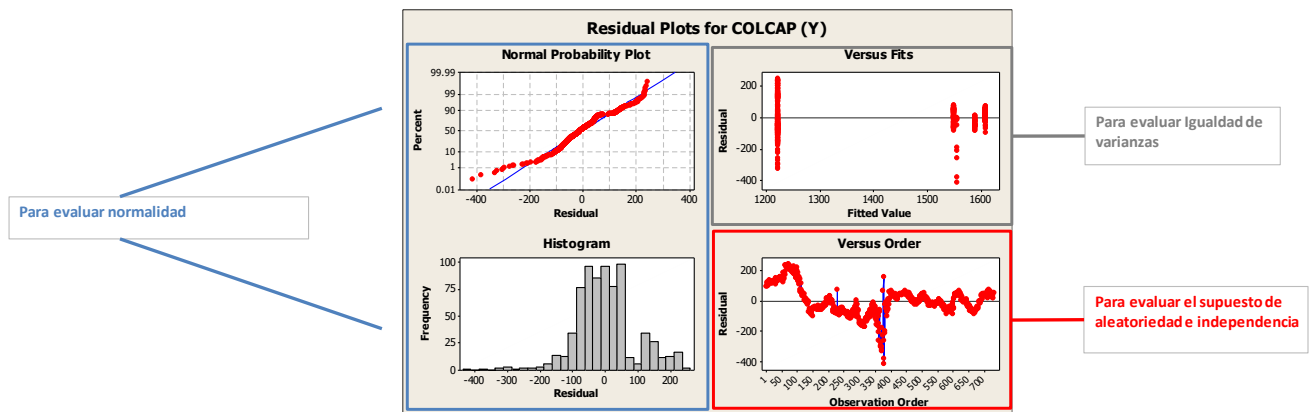


Fig. 193. Gráfico 4 en 1 para validación de supuestos de la tasa prime

El 4 en 1 nos muestra clara cuál es el comportamiento de los datos y ayuda a evaluar e identificar de forma gráfica los supuestos del ANOVA.

- i. Planteamiento de hipótesis para verificar el cumplimiento del supuesto de normalidad de los residuos.

Planteamiento de hipótesis de prueba de bondad de ajuste a la distribución normal

H₀: Los residuos se ajustan a un comportamiento normal

H_a: Los residuos no se ajustan a un comportamiento normal

Regla de decisión: Rechazo H_0 si $p\text{-valor} < \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el $P\text{-valor} = 0.005 < \alpha = 0.05$ **Rechazo H_0**

Conclusión: Con un nivel de confianza del 95%, se puede afirmar que los residuos NO se ajustan a un comportamiento normal. Lo que quiere decir, que hubo factores de ruido que afectaron la información tomada para el experimento y que dañaron el comportamiento normal en las variables.

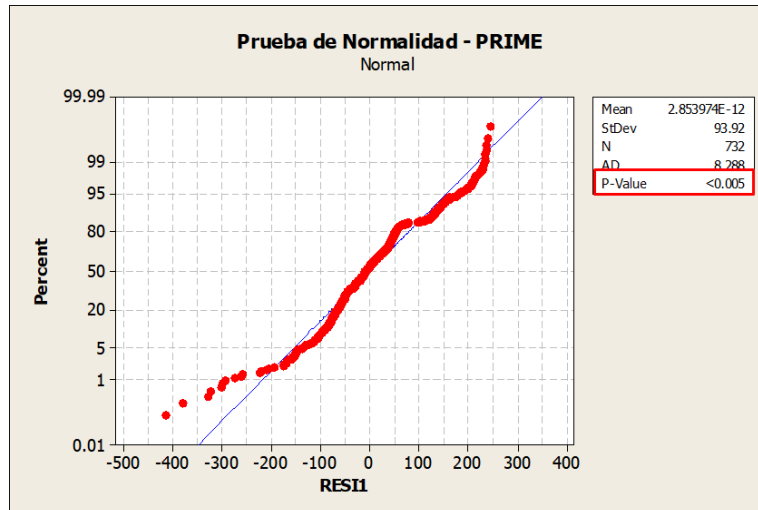


Fig. 194. Prueba de normalidad para los residuos de la tasa prime

j. Planteamiento de hipótesis para verificar el cumplimiento del supuesto de igualdad de varianzas.

Planteamiento de hipótesis de prueba de igualdad de varianzas "homocedasticidad"

Ho: Los residuos se tienen varianzas estadísticamente iguales

Ha: Los residuos no tienen varianzas estadísticamente iguales

Regla de decisión: Rechazo Ho si p-valor $< \alpha$ (para un nivel de confianza del 95%)

Decisión: Como el P-valor = 0.000 $< \alpha = 0.05$ **Rechazo Ho**

Conclusión: Con un nivel de confianza del 95%, se puede afirmar que los residuos NO tienen varianzas estadísticamente iguales. Esto significa que las parejas de datos no oscilan aleatoriamente por encima y por debajo de cero con una variación estadísticamente igual, lo que permite decir que los datos no son comparables entre ellos y hay variables que tienen variación significativamente diferente.

```
Bartlett's Test (Normal Distribution)
Test statistic = 527.56; p-value = 0.000

Levene's Test (Any Continuous Distribution)
Test statistic = 50.66; p-value = 0.000
```

Fig. 195. P-Valor de la varianza obtenido en minitab para la tasa prime

k. Planteamiento de hipótesis para verificar el cumplimiento del supuesto de aleatoriedad e independencia de los residuos.

Planteamiento de hipótesis de prueba de aleatoriedad e independencia

Ho: Los residuos son aleatorios e independientes

Ha: Los residuos no son aleatorios ni independientes

Regla de decisión:

Si D.W. $< dl$ Rechazo Ho

Si $dl < D.W. < du$ Cumple con reserva

Si $D.W. > d_u$ No rechazo H_0

Decisión: Como $D.W. = 0.138886 < d_l = 1.748$ **Rechazo H_0**

Conclusión: Con un nivel de confianza del 95%, podemos afirmar que los residuos NO son aleatorios e independientes, lo que quiere decir que si existe autocorrelación y sesgos entre las variables analizadas.

Datos	
n =	732
k =	2
Nivel de confianza =	95%
d_l =	1.748
d_u =	1.789

Fig. 196. Datos para encontrar el valor d_l y d_u en tablas para la tasa prime

Dato arrojado en Minitab
Durbin-Watson statistic = 0.138886

Fig. 197. Prueba de aleatoriedad e independencia para la tasa prime

l. Conclusión

Después de plantear y validar las respectivas hipótesis para el supuesto de normalidad, igualdad de varianzas y aleatoriedad e independencia a un nivel de confianza del 95%, se puede decir que el modelo no es confiable para realizar estimaciones del índice COLCAP según la tasa prime a corto y largo plazo, ya que los 3 supuestos del ANOVA no se cumplen.

Por otra parte, con base al coeficiente de determinación (R^2 ajustado) el modelo mostró que el 78.87% de la variación del índice COLCAP está siendo explicado por él, y el 21.13% restante está explicado por otros factores que no han sido tomados en cuenta en el modelo.