

# TRABAJO APLICADO DE ESTADÍSTICA



Alumno: Voegeli Walter.

Primera entrega: Análisis descriptivo  
Regresión y correlación



---

## Índice

---

<i>1- Introducción al tema</i>	<i>Pág 2</i>
<i>2- Análisis descriptivo y exploratorio de datos de Bitcoin</i>	<i>Pág 3</i>
<i>3- Regresión y correlación</i>	<i>Pág 6</i>
<i>4- Conclusiones</i>	<i>Pág 10</i>

## Introducción al tema

Las criptomonedas son un tipo de divisa que utiliza un cifrado digital en sus operaciones, como las transferencias de dinero o el pago de un determinado producto o servicio. A diferencia de las monedas tradicionales, solo existe en Internet, y su generación y almacenamiento es puramente digital.

Bitcoin (BTC), nacida en 2008, es la primera criptomoneda (A la derecha su logo). Se creó con el objetivo de proporcionar a los ciudadanos un medio de pago que posibilite la ejecución de transferencias de valor de forma rápida, a bajo costo y que, además, no pueda ser controlado ni manipulado por gobiernos, bancos centrales o entidades financieras. El valor de las criptomonedas se basa puramente en la ley de oferta y demanda, por lo que muchas personas lo utilizan como medios de inversión y especulación aparte de realizar pagos por bienes o servicios.



Al ser bitcoin la moneda con mayor popularidad y madre del concepto criptomoneda, el desempeño de la misma en el mercado influye fuertemente en la psicología de los comerciantes de estas monedas, impactando en el precio de todas las demás criptomonedas alternativas a bitcoin.

A continuación se pretende analizar distintas estadísticas, cómo evolucionó bitcoin, qué relaciones hay entre ellas y cómo podemos estimar a futuro el valor.

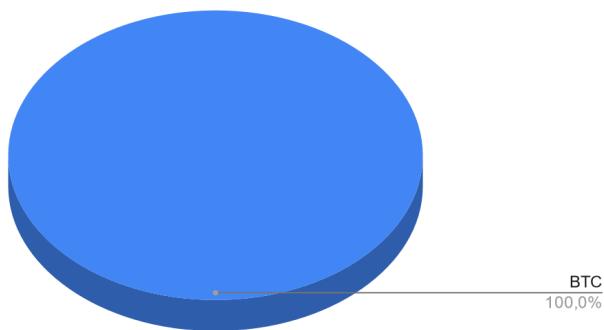
Bien la intro

## Análisis descriptivo y exploratorio de datos de Bitcoin

Se comenzará con el análisis de la capitalización del mercado cripto, esta es calculada multiplicando el precio del momento por la cantidad de unidades que hay en circulación, expresado en miles de dólares.

cap. mercado - Enero 2010

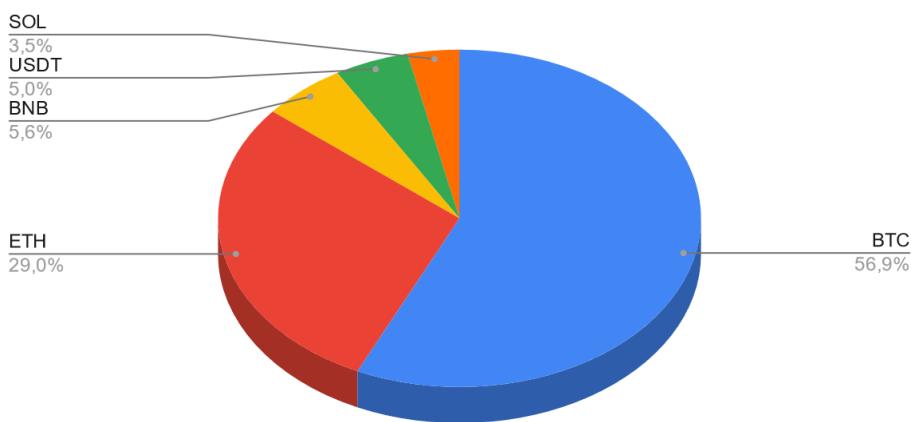
No tiene mucho sentido este gráfico



**FIG 1:** En enero del 2010 BTC poseía el 100% del mercado cripto, con una capitalización de 177 mil dólares.

Al principio bitcoin dominaba el completo mercado cripto, luego en base a esta nueva idea, nacen otras criptomonedas que prometen y aportan diferentes funcionalidades y, debido a esto, el mercado está mucho más diversificado. Aún así, bitcoin sigue siendo la principal moneda.

Top 5 cap. mercado - Enero 2022



**FIG 2:** En enero del 2022, BTC poseía un 40% del mercado cripto (57% en top 5), con una capitalización de 895 mil millones de dólares.

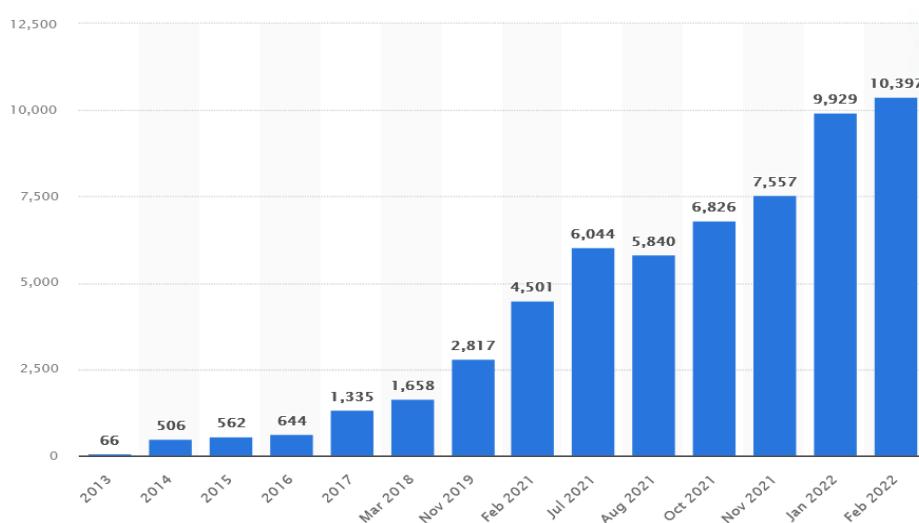
En la FIG 2 se puede observar las principales 5 criptos que más capitalización tienen, siendo BTC la moneda dominante, en segundo lugar ETH (ethereum) nacida en 2015 y seguido de BNB (binance smart chain) nacida a finales de 2020.

Como vimos, con la existencia de otras monedas la dominancia de bitcoin en el mercado cayó, dando lugar a otras monedas alternativas a esta. En el siguiente gráfico podemos observar la dominancia de bitcoin en el mercado a través de los años.



**FIG 3-A:** dominancia de BTC en el mercado cripto entre un rango de 0 a 100 por ciento.

Observando la figura 3, notamos como BTC pasó de ser el casi 100% dominante del mercado en 2016 a ocupar un 40% en enero de 2022. Esto significa que muchas más personas prefieren comercializar con altcoins (monedas alternativas a btc) que con el propio btc a la vez que hay más opciones disponibles.



**FIG 3-B:** Número de criptomonedas distintas que existen en el mercado a través de los años.

En la FIG 3-B, tenemos un gráfico de barras acumulativo que nos indica cuántas criptomonedas aproximadamente existen en distintas fechas, claramente este valor crece y se acumula con el pasar del tiempo, dando más opciones a elegir a los poseedores de criptomonedas.

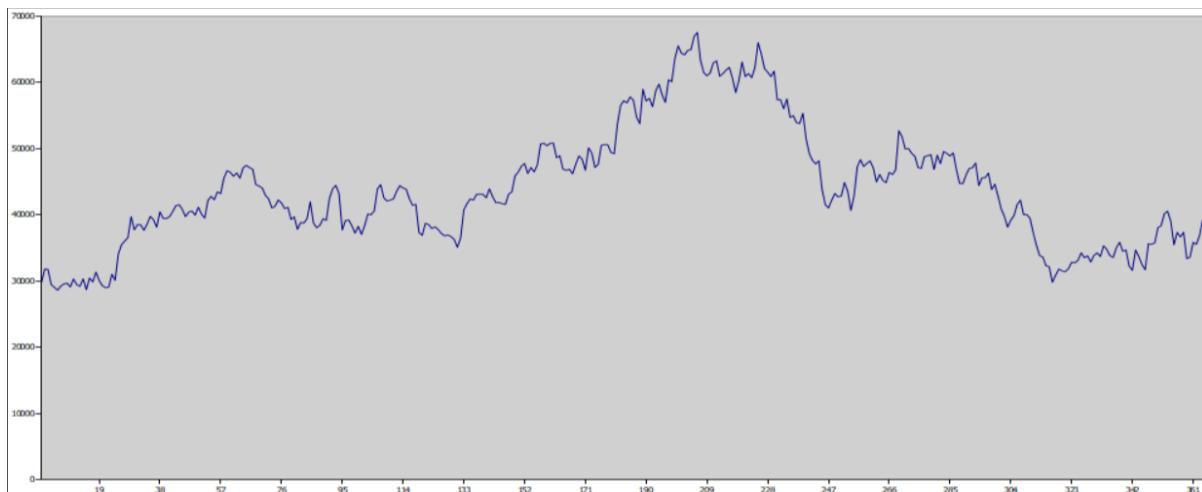
## Regresión y correlación



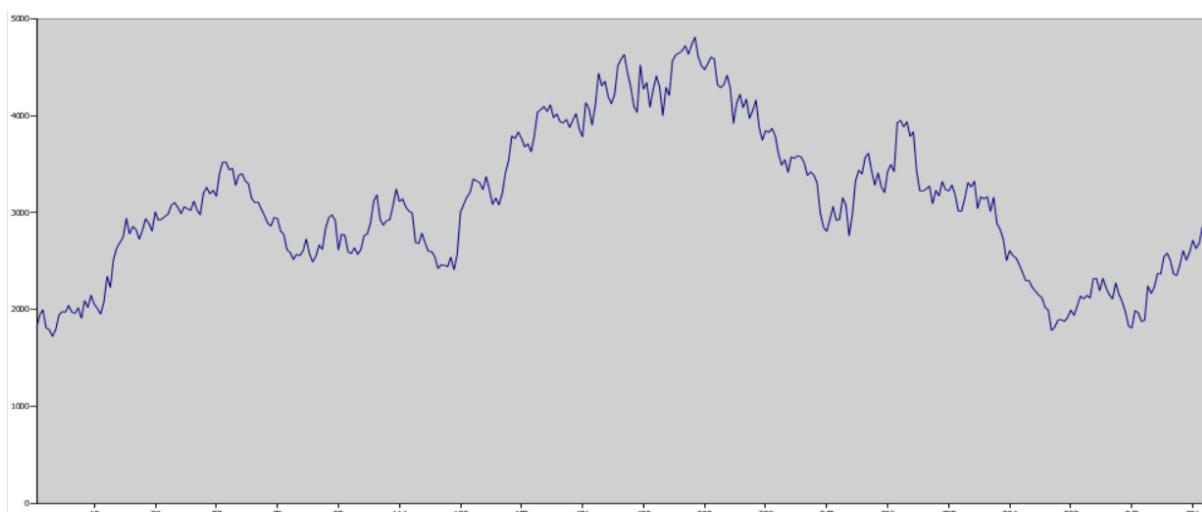
Una moneda alternativa de las miles que existen es Ether o también mal llamada Ethereum (ETH), esta es la segunda más importante y famosa luego de bitcoin.  
(Izquierda: logo de ETH)

Teniendo en cuenta que bitcoin es la número 1 y es la más famosa ¿podemos afirmar que su valor influye en las demás criptomonedas? En resumidas cuentas es un si, el desempeño de esta variable afecta notablemente a las alt-coins. Si bitcoin ingresa en un mercado bajista (que disminuye su precio), las alt-coins también lo harán y viceversa, con mayor o menor retraso, no importa su fundamento.

Teniendo en cuenta lo mencionado y por motivos prácticos, veremos cómo es la relación entre las variables BTC y ETH por días en un periodo de un año.



**FIG 4:** Precio de BTC desde 1 de junio de 2021 a 1 de junio 2022.



**FIG 5:** Precio de ETH desde 1 de junio 2021 a 1 de junio 2022.

A priori hay una notable semejanza entre ambas variables ¿qué nos dice el análisis de relación?

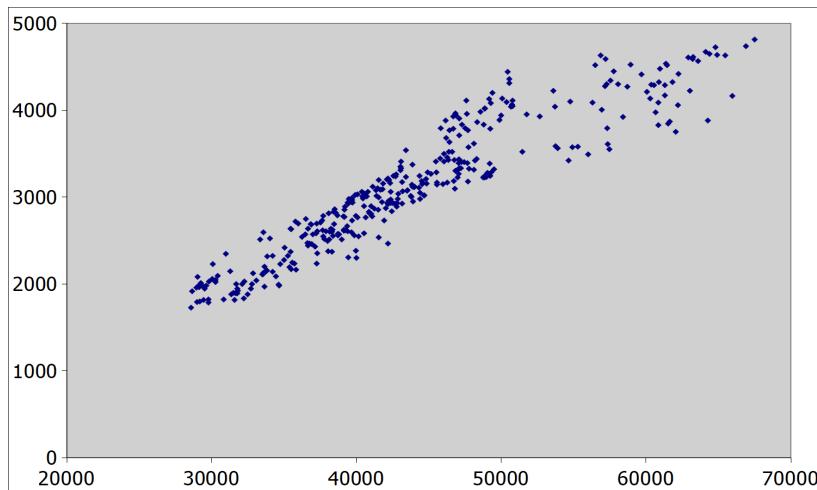
### Gráfico de dispersión.

El dispersiograma es una gráfica de puntos dispersos que muestra la variabilidad o dispersión existente entre ambas variables. Tomaremos como variable independiente a BTC ya que rige el precio y dependiente a ETH.

Para la construcción de la gráfica utilizaremos:

Eje X - histórico diario del precio de cierre diario de bitcoin.

Eje Y - histórico diario del precio de cierre diario de ethereum.



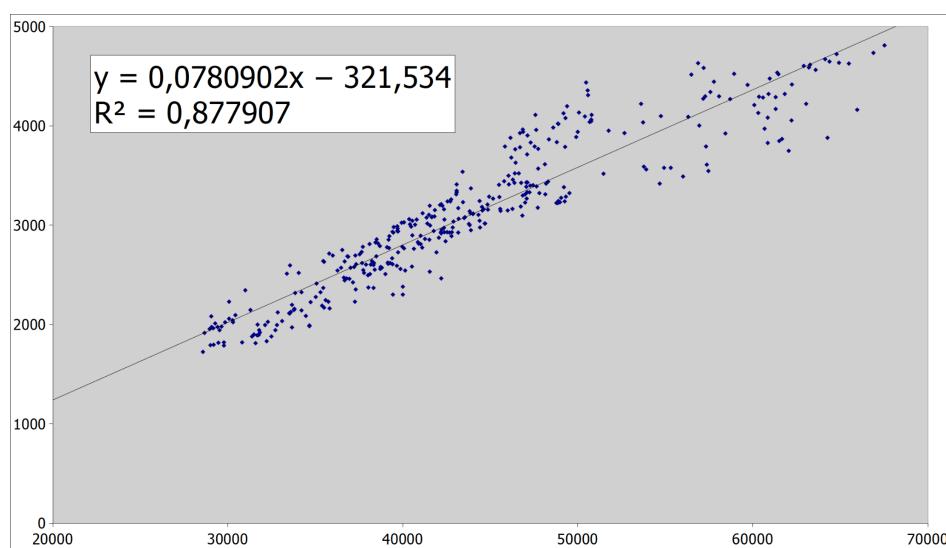
**FIG 6:** dispersiograma precio BTC- precio ETH.

### Modelo

#### Recta de regresión.

El análisis de regresión estudia la relación en la que interviene una variable independiente (precio btc) y una variable que asumimos dependiente (precio eth), veremos si la relación entre ellos es lineal.

- $\alpha$  : intersección con eje y = -321,53
- $\beta$  : pendiente = 0,078
- (la pendiente es la proporción de cambio medio de ETH por cada cambio de BTC)
- Ecuación de regresión  $E(y) = 0,078x - 321,53$



**FIG 7:** La relación entre ambas monedas se puede aproximar mediante una línea recta ascendente.

A continuación se procedió a analizar la regresión y correlación con datos obtenidos del software Gnumeric.

SALIDA RESUMEN	
Estadísticas de regresión	
R múltiple	0,936967
R <sup>2</sup>	0,877907
Error estándar	263,532329
R <sup>2</sup> ajustado	0,877572
Observaciones	366,000000

**FIG 8:** Datos obtenidos por el software.

- **Coeficiente de correlación (R):** índice que puede utilizarse para medir el grado de relación lineal de dos variables siempre y cuando ambas sean cuantitativas y continuas.

$$R \text{ obtenido} = 0.937$$

Al ser R aproximado a 1, entonces hay un grado de relación lineal muy bueno entre ambas variables.

- **Coeficiente de determinación (R<sup>2</sup>):** Refleja que tan bien se ajusta el modelo adoptado a la variable real.

$$R^2 = 0.878$$

Esto significa que nuestro modelo representa un 87.8% a lo que estimamos, lo cual indica el porcentaje de veces BTH influencia en el precio de ETH en el periodo analizado.

Se concluye que la dependencia entre ambas variables se aproxima a la linealidad.

No está del todo bien este concepto.  
Lo que indica es el porcentaje de causas comunes entre las variables relacionadas.

### Análisis de residuos.

El residuo es la diferencia entre el valor observado (puntos en la FIG 7) de la variable dependiente ( $y_i$ ) y el valor estimado del modelo de ella (Recta FIG 7).

Residual
-187,3684284536953
-219,43066558501187
-157,28653915115274
-165,85094856203773
-152,5196106066362
-187,6409736853734
-165,55473047728879
-42,388953314005676
-14,390509198458858
22,560814929433946
-0,3642788440884033
-1,90138695182236
2,8197167153373357
-25,72142064399577
-6,263419203110516

**Fig 9:** Primeros 15 valores de residuos de los 366 obtenidos.

- Los supuestos teóricos sobre los que se sustenta el Modelo De Regresión **Lineal Simple** son 4 :

**1. LINEALIDAD:** La relación entre las variables debe ser LINEAL, esto ya lo comprobamos viendo en la figura 7 que se ajusta muy bien el modelo lineal a nuestro caso.

**2. INDEPENDENCIA:** Los errores son Independientes

**3. HOMOCEDASTICIDAD:** Los errores tienen una varianza constante

**4. NORMALIDAD:** Los errores están distribuidos normalmente

Entonces ahora tenemos que comprobar los supuestos 1, 2 y 3 en nuestros valores residuales.

### Lectura de valores residuales

Valor del residuo Mínimo: -821,21

Valor del residuo Máximo: 815,75

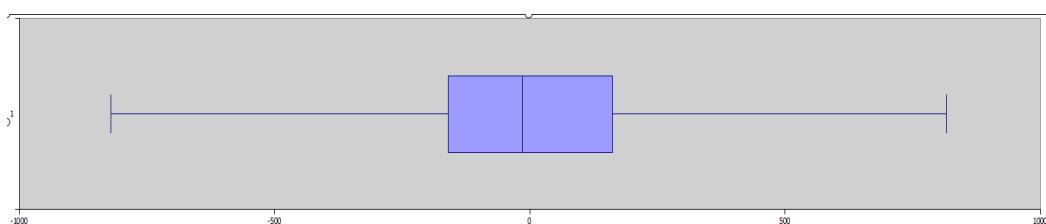
Rango: 1636,96

Mediana: -14,58

### Gráfico Box-Plot

El gráfico boxplot muestra un resumen de una gran cantidad de datos (residuos en este caso) obtenidos en 5 medidas, facilitando la comprensión del mismo.

Los extremos son los valores máximos y mínimos, la caja es el rango donde mayor cantidad de valores se encuentran y por último la mediana, que divide la caja en dos y nos ayuda a ver su simetría.



**FIG 10:** Gráfico box-plot de los residuos obtenidos.

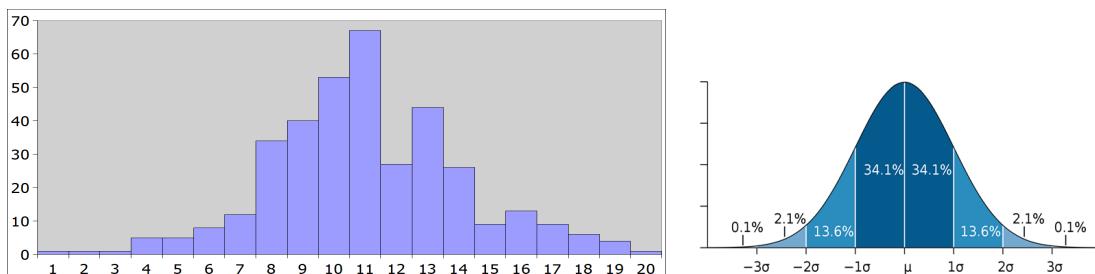
Observando la figura 10, se concluye que la distribución de los residuos es casi simétrica, reforzando que el modelo de análisis aplicado es el indicado.

Al ser una caja chica con respecto al máximo y mínimo, no hay mucha variabilidad de residuos, es decir que se concentran cercanos a la recta lineal de la figura 7.

## Histograma

En este gráfico veremos la frecuencia de apariciones de valores contiguos dentro de intervalos llamados clases.

En nuestro caso, los residuos se agrupan por proximidad en 20 clases, es decir que las clases más altas van a agrupar residuos que aparecerán con mayor frecuencia.

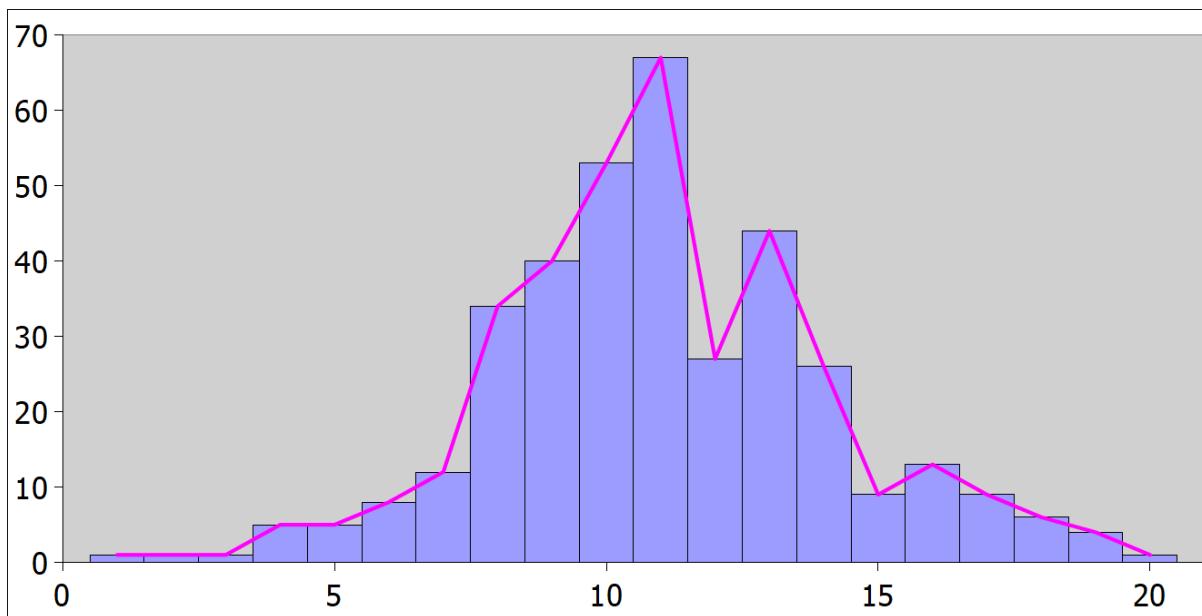


**FIG 11:** (IZQ) Histograma de residuos obtenidos, se asemeja a una campana de gauss (Der).

Podemos concluir que los residuos que se encuentran en la clase 11 son los que más cantidad de veces se repiten. A su vez, también nos confirma el supuesto 4 de normalidad al asemejarse el histograma a una distribución normal o distribución gauss (fig 11 der)

## Polígono de frecuencias.

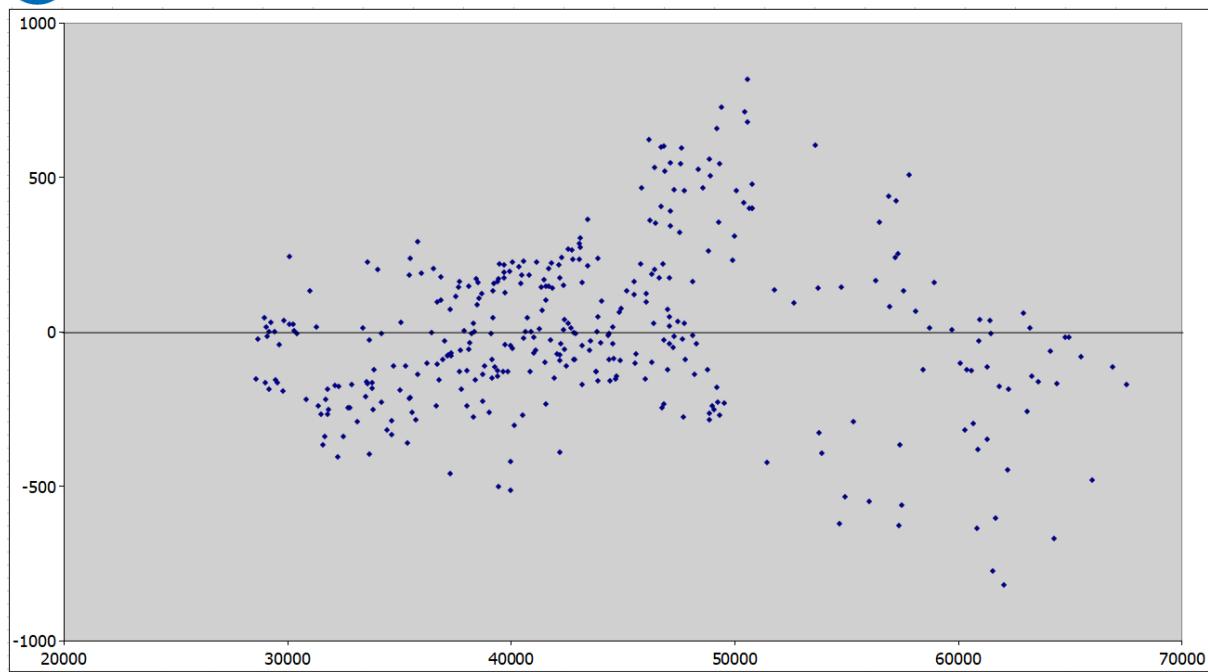
Se construye mediante la unión por una recta de los valores máximos de las clases de un histograma, es otra forma de graficar las frecuencias.



**FIG 12:** polígono de frecuencias sobre el histograma de residuos.

## Gráfico de dispersión de residuos.

Para cada residuo obtenido se grafica un punto denotando su valor en el eje Y correspondiente al valor de BTC en el eje X.



**FIG 13:** Gráfico de dispersión  $r_i$  (residuos) versus  $X_i$  (BTC), se asemeja a una nube de puntos.

La figura 13 nos muestra como el error  $\epsilon$  (eje Y) se mantiene mayormente variable para cada variable X, esto nos confirma el supuesto 2 de independencia ya que los errores no siguen un patrón específico.

También nos confirma el supuesto 3 de HOMOCEDASTICIDAD, ya que se considera que los residuos presentan una distribución homogénea en la figura 3.

Aún así hay cierta pérdida de eficiencia en el modelo lineal en el rango de precio de BTC que va de 50000 a 70000, esto es debido a que si nos fijamos en la figura 4, 5 y 7 hubo una variación considerable de relación entre ambas variables.

Muy bien y completo el análisis de regresión. Se podría realizar en un trabajo futuro alguna predicción con su correspondiente intervalo. Por otro lado, tener en cuenta que los datos sirven para realizar análisis de series de tiempo, que también se podría llevar a cabo en un trabajo futuro.

### Conclusiones

El análisis exploratorio de datos es una práctica importante en el mundo económico, nos permite saber en qué estamos depositando nuestro dinero y a saber por qué suceden ciertos acontecimientos, realizar estadísticas y trabajar en función de ellas. En este caso pudimos ver que la dominancia de bitcoin disminuyó drásticamente al introducirse al mercado nuevas criptomonedas. Luego se analizó qué relación hay entre la principal y segunda moneda líder en el mercado, denotando que entre ellas existe una relación lineal. A pesar de una pequeña discrepancia en el análisis de residuos al no amoldarse perfectamente a los supuestos, se puede asegurar que el modelo de regresión lineal es el más adecuado para representar la relación entre bitcoin y ethereum por cumplir en mayor parte los requisitos de validación del modelo de regresión lineal (pág 7). Gracias a este análisis, uno puede tener el concepto de que cuando el valor de la variable dependiente BTC crezca o disminuya, también lo hará ETH siguiendo la dirección de crecimiento de BTC.

El trabajo está muy completo y se evidencia un claro entendimiento de los temas desarrollados.

Nota: 9

Corregido por María José