Fecha de nacimiento = 11/07/1983

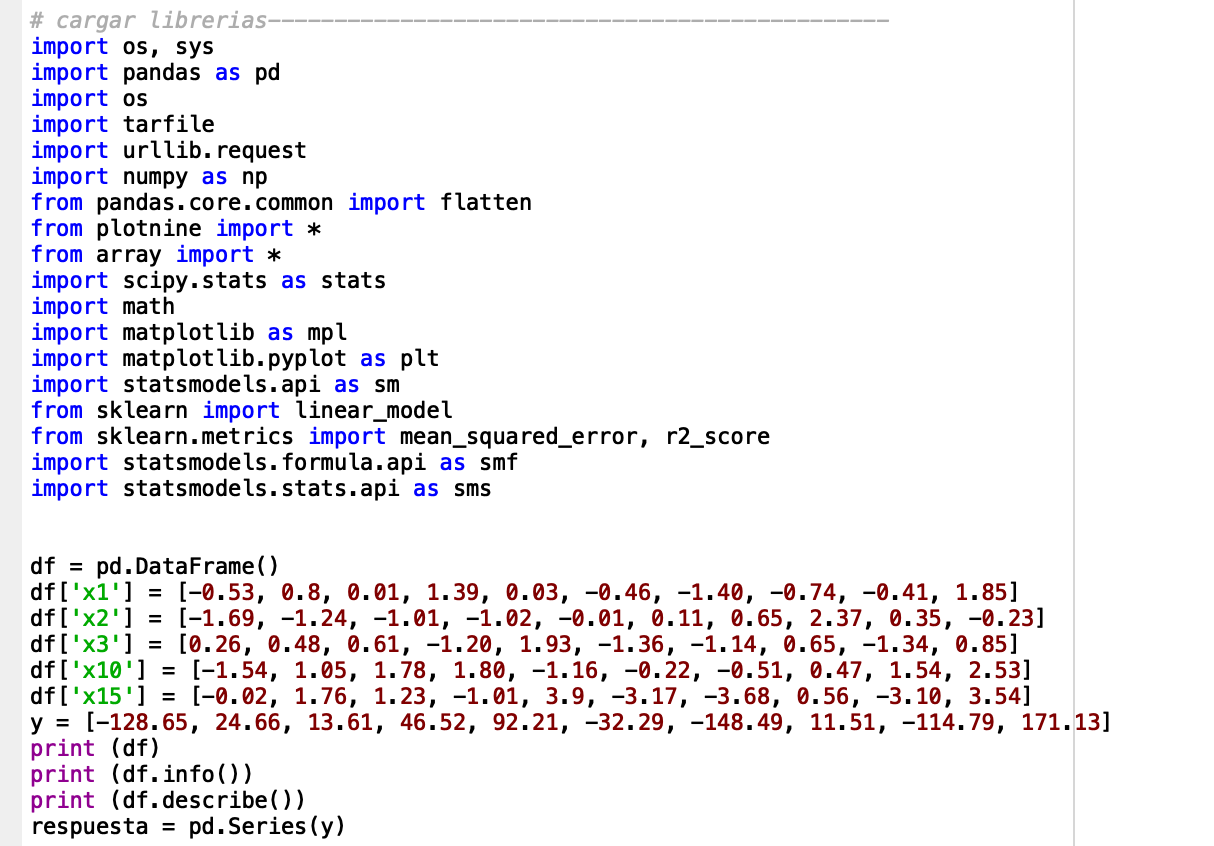
m=10

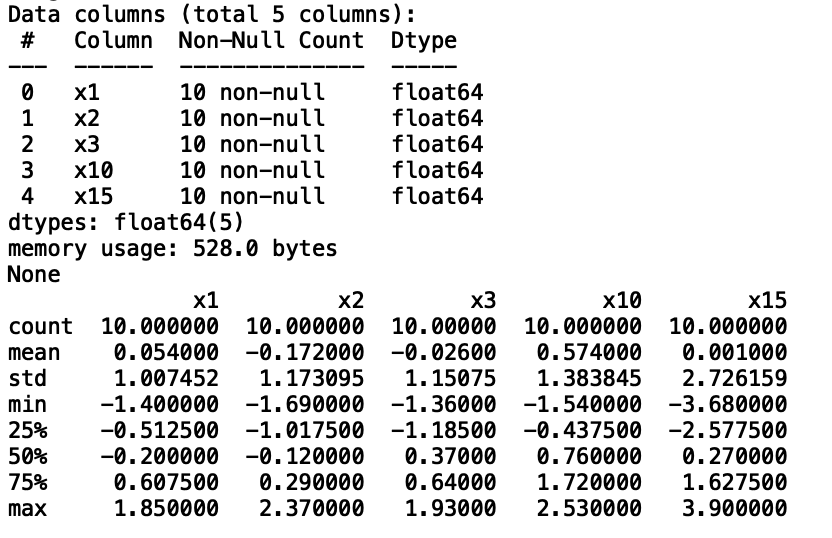
d=15

x1,x2,x3,x10,x15,y

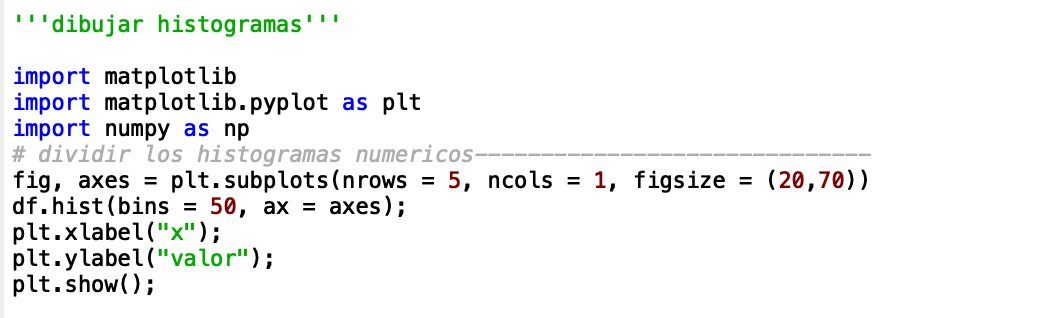
1. **REPRESENTAR LOS DATOS**

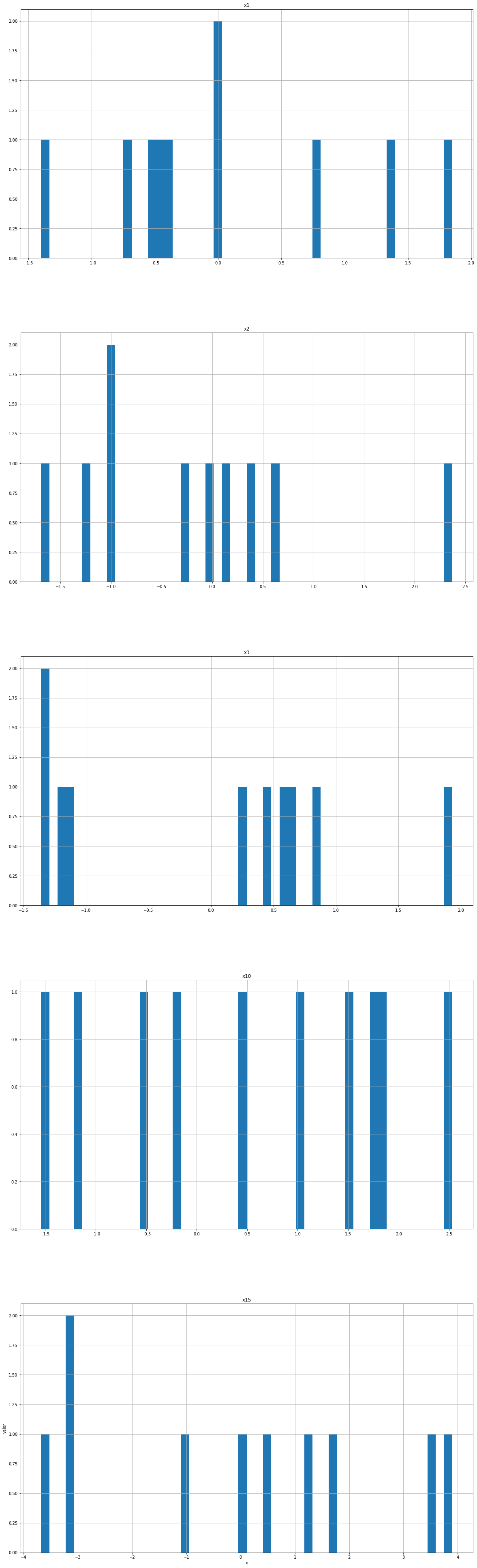
Se cargan librerías, se crea los dataframe





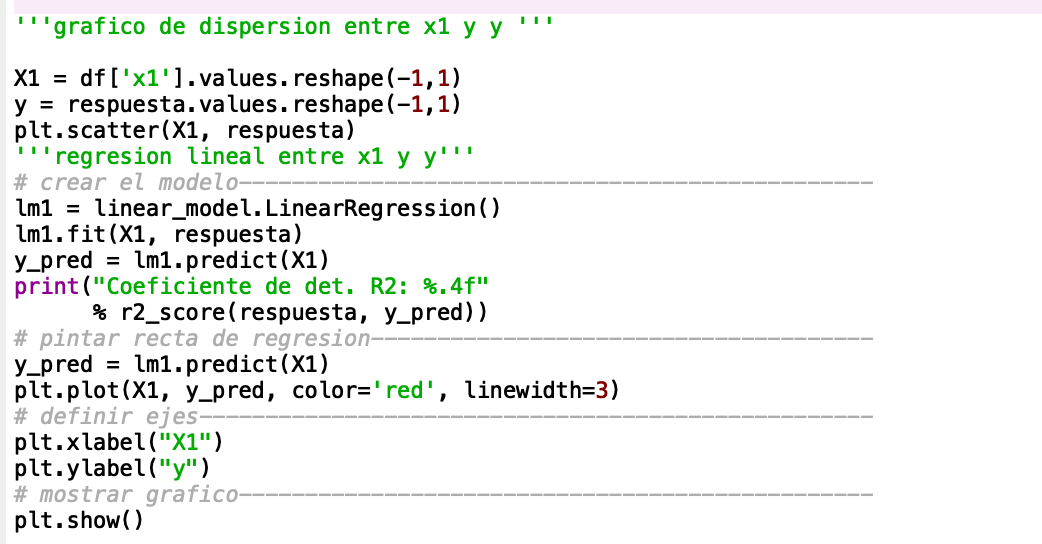
Se dibuja los histogramas de cada variable en un solo gráfico

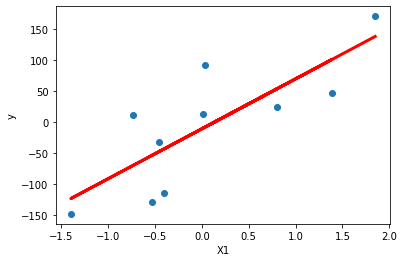


****

**Los histogramas muestran que no existe una normalidad entre los datos.**

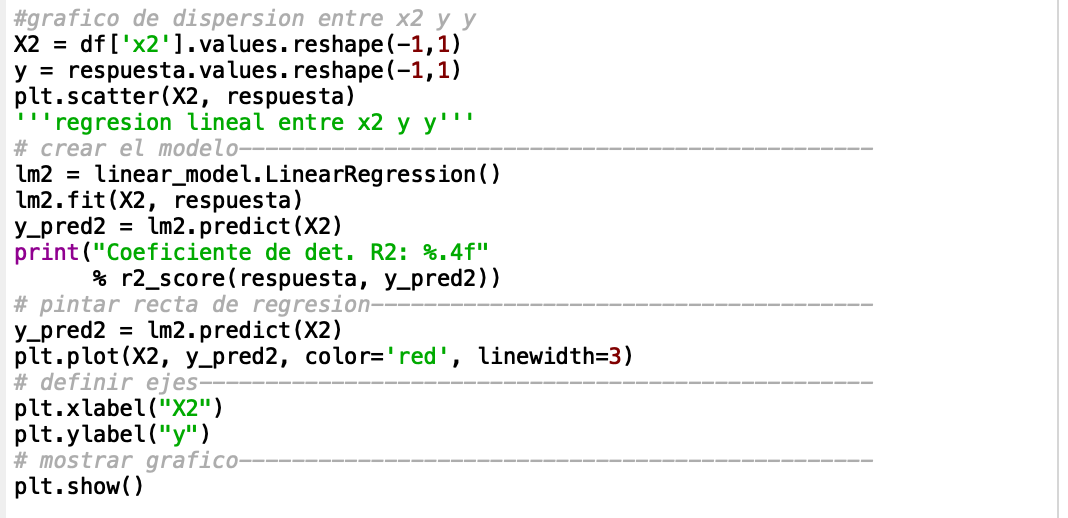
Se procede a graficar cada una de las variables con respecto a “y”

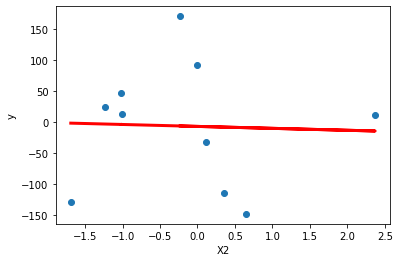






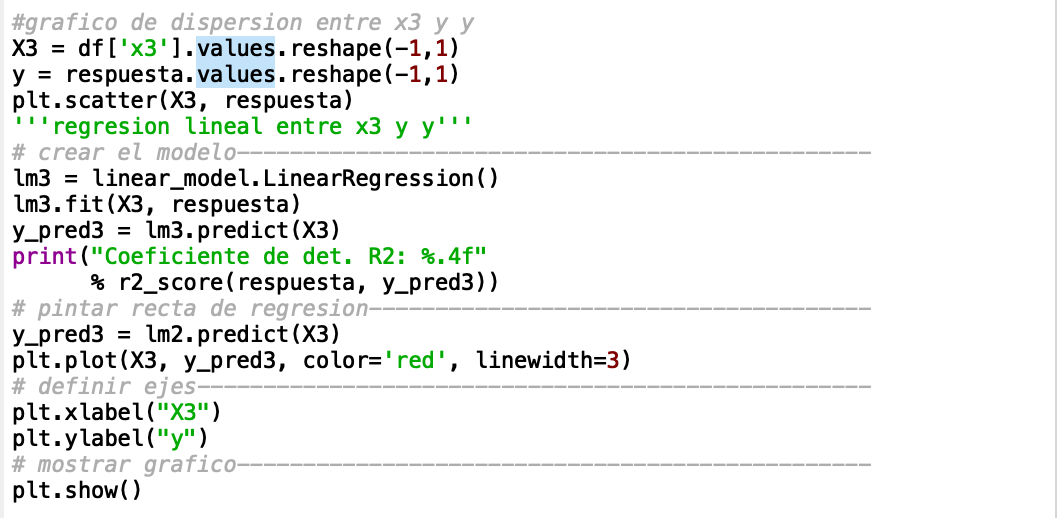
Se observa en la figura que X1 se ajusta bien ya que existe una concentración de la variable alrededor de la bisectriz sin embargo el r2 que nos da la bondad del modelo debería estar por encima del 0.7

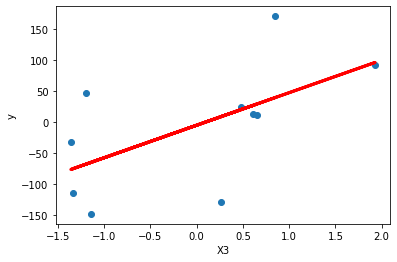






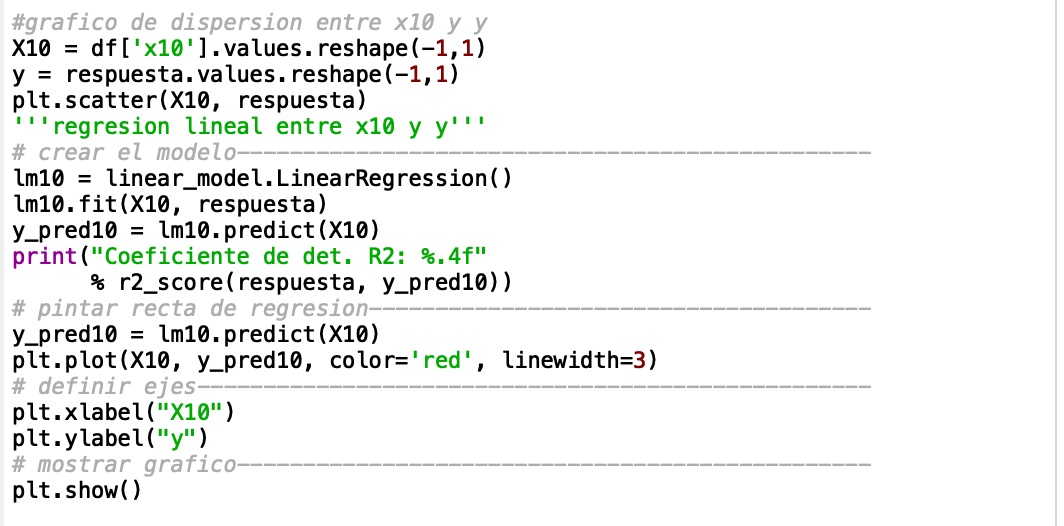
Se observa en la figura que X2 no se ajusta, no existe una concentración de la variable alrededor de la bisectriz y esto se ve reflejado en el r2 obtenido que es muy bajo,.

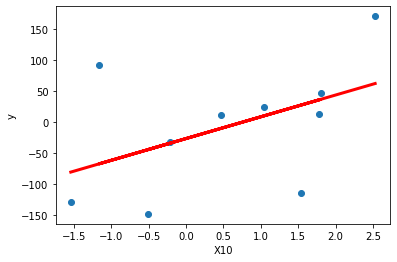






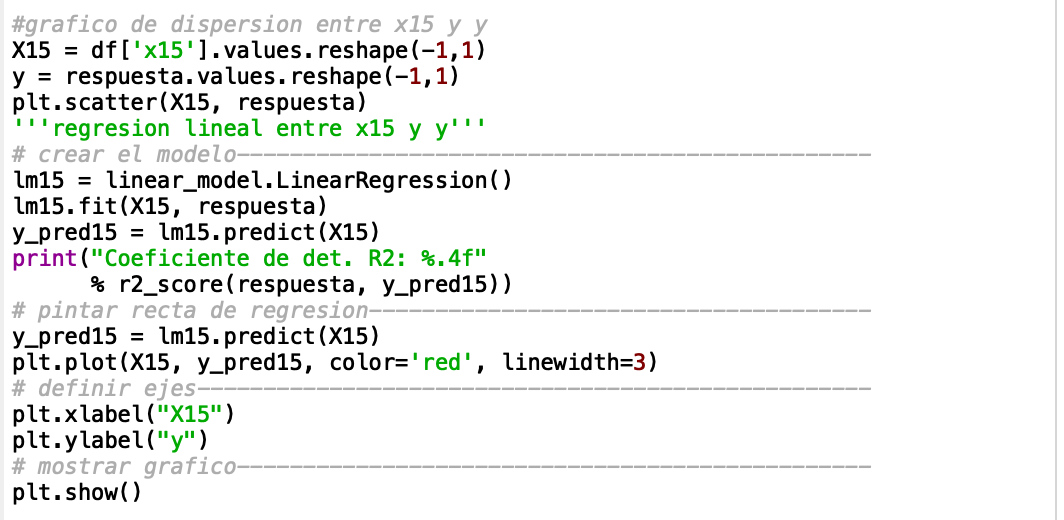
Se observa en la figura que X3 no se ajusta bien ya que no existe una concentración de la variable alrededor de la bisectriz, su r2 también está muy por debajo del deseado. Se ve presencia de posibles outliers.

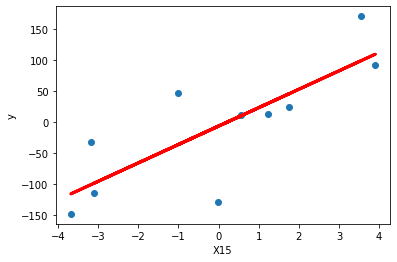






Se observa en la figura que X10 no se ajusta bien ya que no existe una concentración de la variable alrededor de la bisectriz, su r2 también está muy por debajo del deseado. Presencia de outliers.

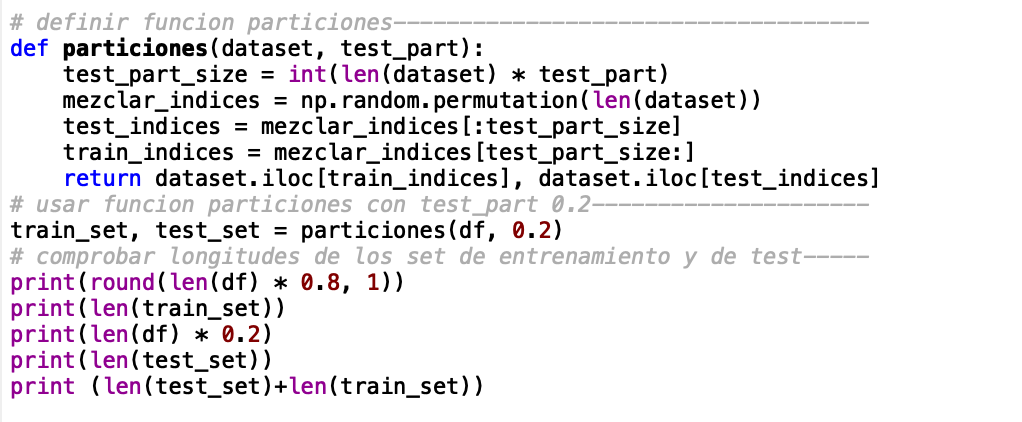






Se observa en la figura que X15 no se ajusta bien ya que no existe una concentración de la variable alrededor de la bisectriz, su r2 también está muy por debajo del deseado. Presencia de outliers.

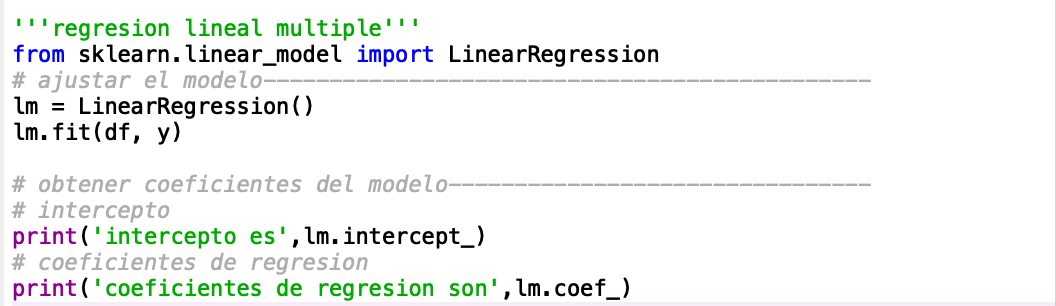
Se procede a dividir el dataset en train y test

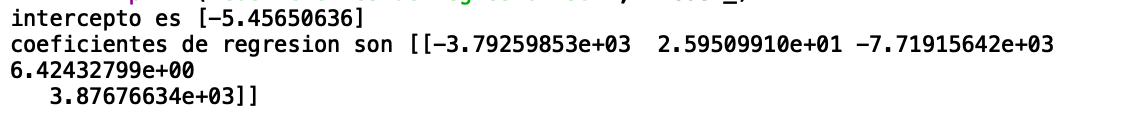


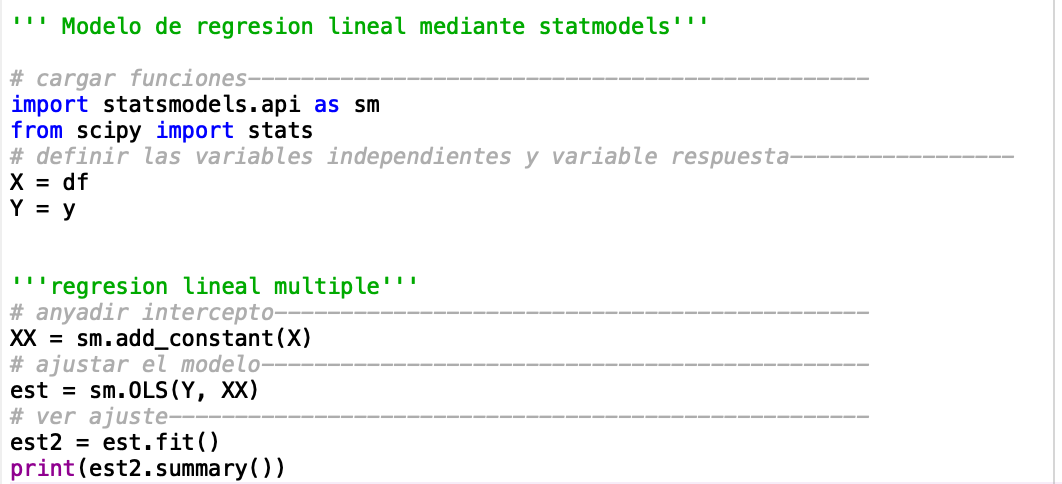
Resultado:

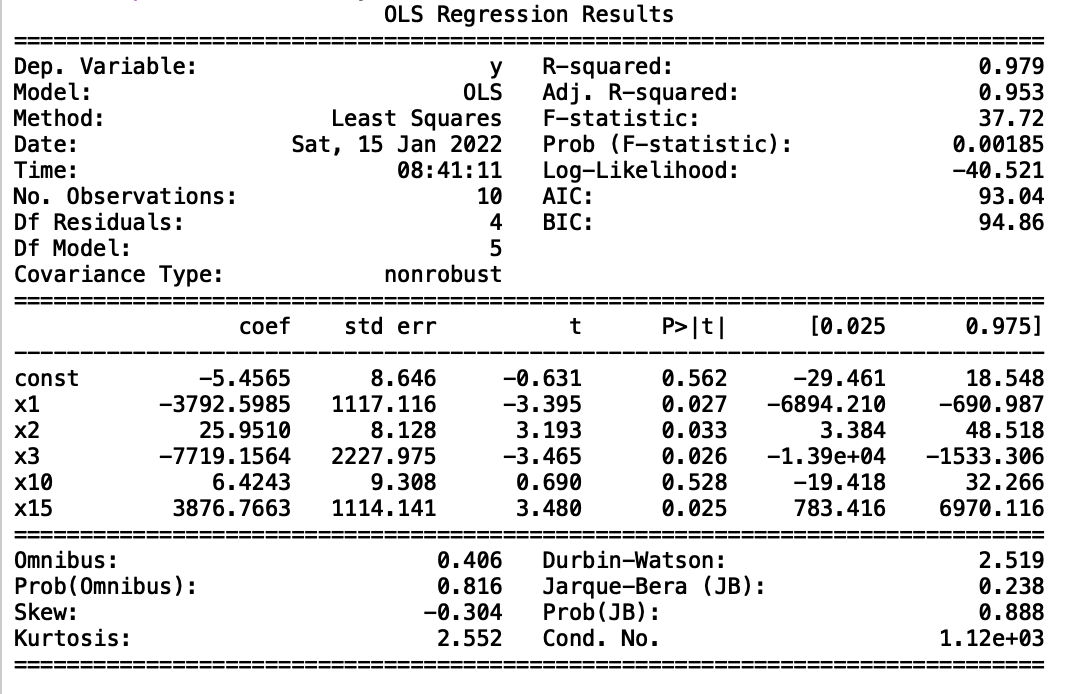


1. **REALIZA UN AJUSTE DE REGRESION LINEAL MULTIVARIANTE**









En la captura anterior, se analiza que se tiene la variable predictora que no son estadísticamente significativas (deben ser su p.valor <0,05), ésta es x10. Además de por el p. valor, se puede comprobar en éstas dos, la no significatividad observando el intervalo de confianza al nivel 1-α donde contienen al 0.

Se realiza el gráfico en el cual se compara los valores reales de la muestra frente a los valores del ajuste.

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

1. ***REALIZA UNA SELECCIÓN DE VARIABLES CON EL METODO STEPWISE***

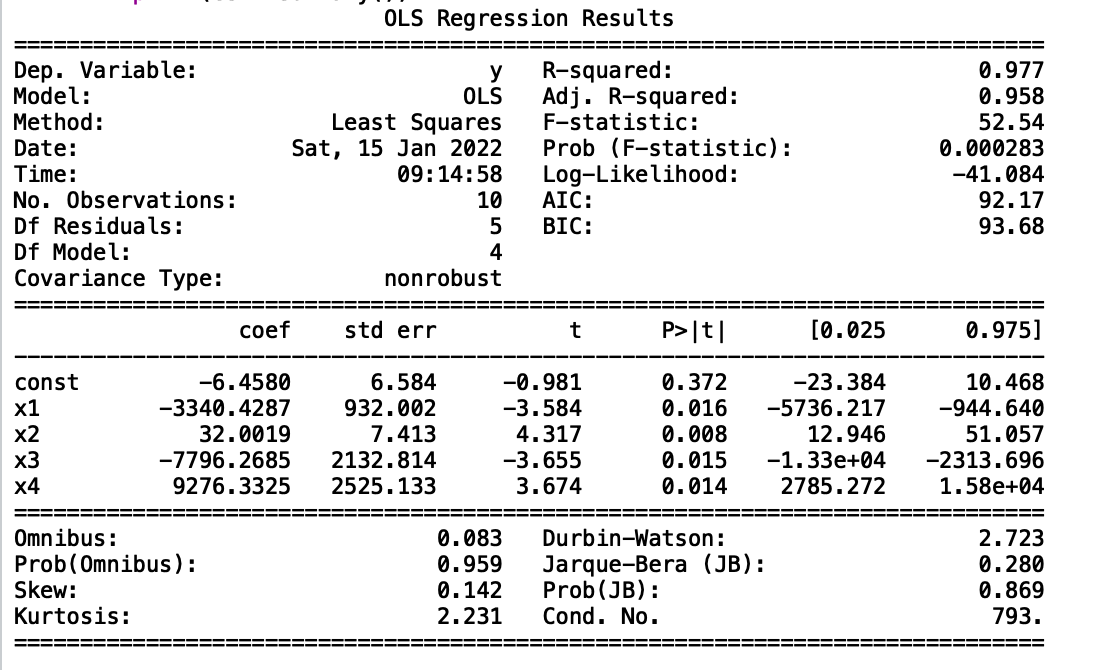
Se usa .drop para eliminar las columnas deseadas. Se inicia con la eliminación de la variable predictora x10 que es la menos significante

Texto

Descripción generada automáticamente

Imagen de la pantalla de un celular de un mensaje en letras negras

Descripción generada automáticamente con confianza baja



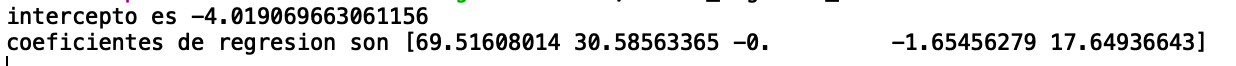
Finalmente, se tiene 4 variables predictoras que son significativas ya que su p.valor es menor a 0.05, esto se ha logrado eliminando X10

Nos quedamos con x1, x2, x3, x15

1. **REGRESION LASSO**

Imagen que contiene Interfaz de usuario gráfica

Descripción generada automáticamente

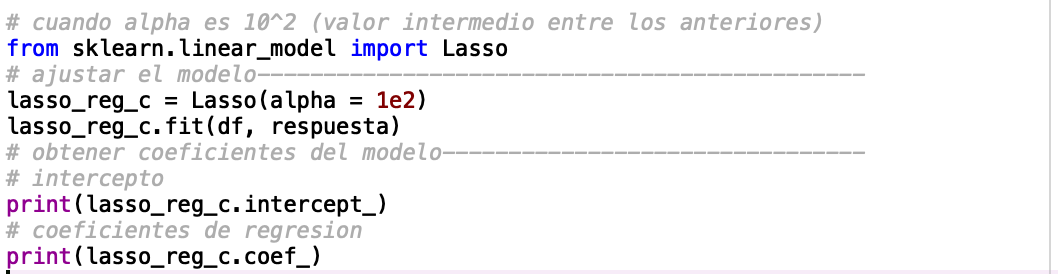


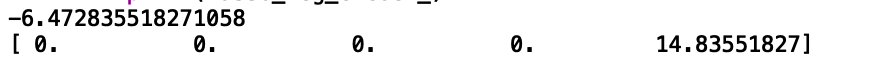
Tabla

Descripción generada automáticamente

Un conjunto de letras negras en un fondo blanco

Descripción generada automáticamente con confianza media





Con alpha 1 no se elimina ninguna variable predictora

Con alpha 10^4 en cambio se elimina todas variables predictoras

Con alpha 10^2 (valor intermedio) se elimina 3 variables predictoras.

1. Compara resultados

**PREGUNTA**

Explica en menos de 200 palabras…..

Conocidos también como outliers, son aquellos que en una gráfica están alejados de la zona donde se encuentran los demás puntos y también están alejados de la recta de regresión.

***¿Que consecuencias tiene?***

La recta de regresión se ve muy afectada por la presencia de estos outliers, su pendiente se ve afectada, es como que estos puntos halaran a la recta hacia ellos. Además, la media aritmética se desplaza hacia el valor máximo de la distribución.

***Como podemos detectarlo?***

Se los detecta mediante técnicas de visualización que permiten comprobar la existencia de datos anómalos.

Un modo intuitivo de comprobar si la muestra puede contener outliers que distorsionen los resultados es calcular estimadores clásicos y robustos y realizar un análisis de sensibilidad analizando las diferencias entre ellos. Si ambos tipos de estimadores difieren poco, probablemente no haya grandes anomalías en la muestra y los estimadores clásicos sean adecuados; en otro caso, puede ser más conveniente utilizar estimadores robustos.

***Como podemos trabajar si tenemos una muestra con datos anómalos?***

1. Detectar los outliers, y eliminarlos de la muestra para, posteriormente, aplicar los métodos tradicionales de estimación (NO RECOMENDABLE PORQUE PODEMOS ESTAR ELIMINANDO INFORMACION VALIOSA).
2. Diseñar métodos robustos de estimación que no se vean muy afectados por la existencia de estos outliers aunque, si la muestra no tuviera valores anómalos, las estimaciones fueran algo menos eficientes que las tradicionales.