**Regresión lineal – Ejercitación – Metodología de Análisis en Opinión Pública**

Nombre del alumno/a: Mercedes Marani, Paula Mendoza, Rocío Nuñez

1. Utilizando la base de datos de una encuesta a Diputados Nacionales en Argentina en el año 2022, realice las siguientes consignas:

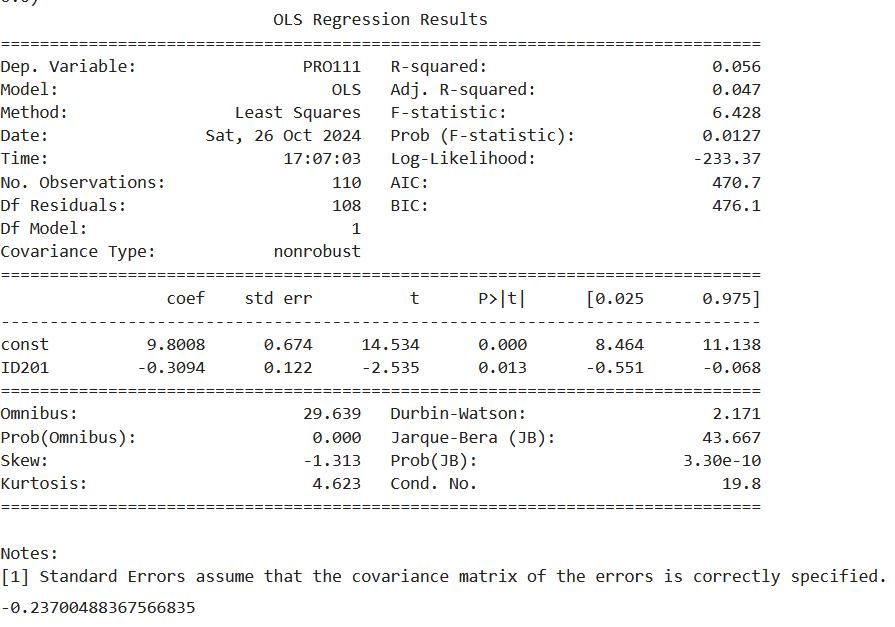
a. Elija distintas variables para hacer tres regresiones lineales simples en Python.

b. Interprete los resultados obtenidos a partir de los distintos elementos de las regresiones, por ejemplo, los coeficientes, el R2, el nivel de confianza/significatividad, etc. ¿A qué conclusiones podríamos llegar a partir de los datos obtenidos?

**Regresión lineal 1:** Opinión sobre el grado de importancia de la deuda externa según la posición ideológica del partido.

**Variable 1:** [ID2] Y, en esta misma escala, ¿dónde situaría Ud. a su propio partido político?

**Variable 2:** [PRO111] Grado de importancia de “Deuda externa”



**Análisis**

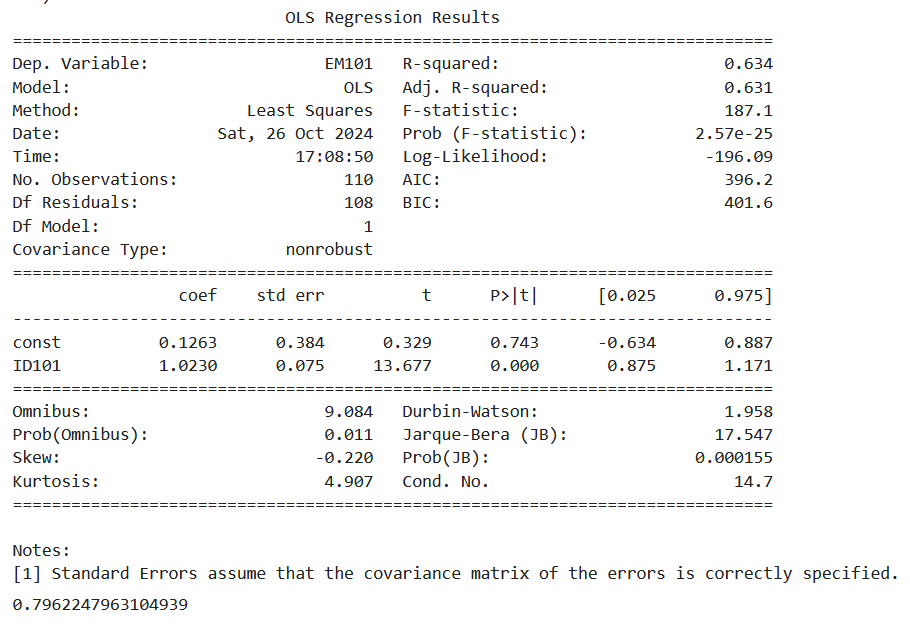
El coeficiente de correlación es -0.23, lo que implica una relación negativa débil. Esto representa que una variable aumenta mientras otra disminuye pero, en este caso, es una correlación baja, ya que la correlación perfecta se encuentra en 1 o -1.

Con respecto al nivel de confianza, tenemos que la variable x “ID2” es -0.3, lo que implicaría que hay una relación inversa. El valor P es 0.013, lo que supone un nivel de confianza de 99%; esto permite entender que hay un alto nivel de confianza en la muestra.

En base al resultado de R-squared de 0.056, podemos deducir que la posición ideológica del partido según la percepción del diputado/a (variable x) no permite explicar, en gran medida, de forma lineal, el grado de importancia asignado a deuda externa (variable y); solo lo explica en un 5%. La conclusión es que la ideología no es un predictor fuerte para explicar esta posición y probablemente deberíamos considerar otros factores.

Podemos decir que si bien la relación entre las variables es estadísticamente significativa (según el valor p), el modelo no es muy útil para predecir la variable dependiente debido al bajo valor de R².

**Regresión lineal 2:** Opinión sobre economía regulada por el Estado o por el mercado en base a auto posición ideológica del diputado/a.

****

**Variable 1:** [ID1] Como recordará, cuando se habla de política se utilizan normalmente las expresiones izquierda y derecha. En esta tarjeta hay una serie de casillas que van de izquierda a derecha ¿En qué casilla se colocaría Ud. teniendo en cuenta sus ideas políticas?

**Variable 2:** [EM1] Dentro del debate económico sobre modelos de regulación. ¿Podría decirme si está Ud. ¿Más a favor de una economía regulada por el Estado o de libre mercado? Utilice para ello la siguiente escala de 1 a 10, donde el "1" significa una máxima regulación del Estado en la economía y el "10" una máxima liberalización de la economía a través del mercado.

**Análisis**

El coeficiente de correlación es 0.79, lo que implica una relación positiva fuerte. Esto representa que mientras una variable aumenta la otra también. En este caso, es una correlación alta, ya que la correlación perfecta se encuentra en 1 o -1.

Con respecto al nivel de confianza, tenemos que la variable x “ID1” tiene un coef: 1, lo que implicaría que hay una relación positiva. El valor P es 0.0, lo que supone un nivel de confianza de 100%, esto permite entender que hay un alto nivel de confianza en la muestra.

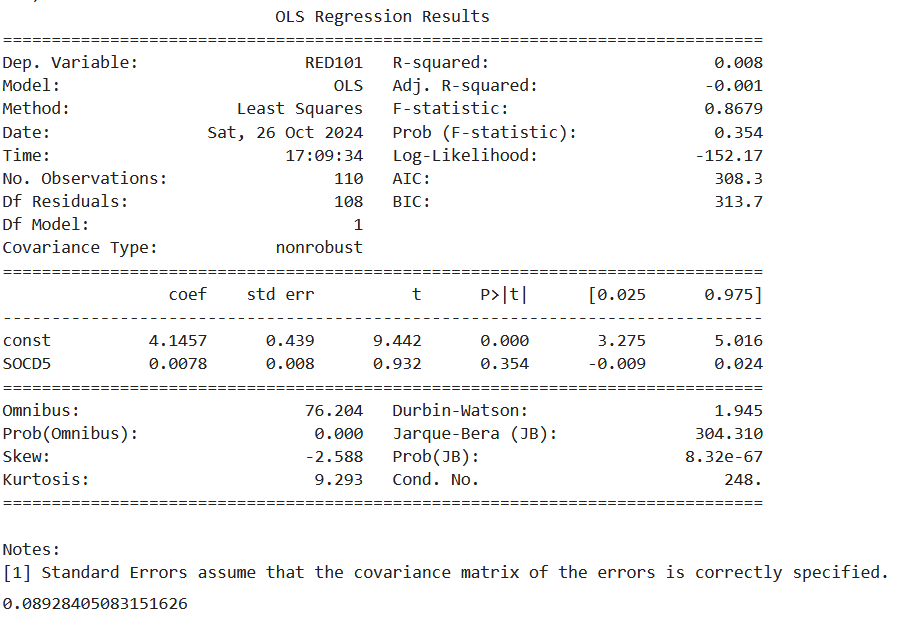
En base al resultado de R-squared de 0.634, podemos deducir que la posición ideológica del diputado/a (variable x) permite explicar en un 63%, de forma lineal, la posición tomada en torno a quien debería administrar la economía (variable y). La conclusión es que la ideología es un predictor fuerte ya que suele estar muy relacionada con las posiciones respecto a los roles del Estado y el libre mercado en la economía.

Podemos decir que la relación entre las variables es estadísticamente significativa (según el valor p), y el modelo es muy útil para predecir la variable dependiente debido al alto valor de R².

**Regresión lineal 3:** Frecuencia de utilización de red social Twitter según edad del diputado/a

**Variable 1:** [SOCD5] ¿Cuántos años cumplió Ud. en su último cumpleaños?

**Variable 2:** [RED101] ¿Con qué frecuencia, diariamente, algunas veces a la semana, algunas veces al mes, algunas veces al año o nunca, utiliza Ud. o alguien de su equipo las siguientes redes sociales para interactuar con los ciudadanos? Twitter.



**Análisis**

El coeficiente de correlación es 0.08, lo que implica una relación positiva débil. Esto representa que mientras una variable aumenta la otra también. Pero, en este caso, es una correlación baja, ya que la correlación perfecta se encuentra en 1 o -1.

Con respecto al nivel de confianza, tenemos que la variable x “SOCD5” tiene un coef: 0.007, lo que implicaría que hay una relación positiva. El valor P es 0.354, lo que implica que el nivel de confianza es 65%. Permitiendo entender que hay un bajo nivel de confianza en la muestra.

En base al resultado de R-squared de 0.008, podemos deducir que la edad (variable x) no permite explicar, de forma lineal, la frecuencia de utilización de la red social Twitter (variable y). La conclusión es que la edad no es un predictor fuerte para deducir la frecuencia de uso de esta red social por los diputados o sus equipos.

Podemos decir que la relación entre las variables no es estadísticamente significativa (según el valor p) y el modelo no es muy útil para predecir la variable dependiente debido al bajo valor de R².