**TP 5: EJERCICIO 3**

use "CEOSAL2.dta"

sum, detail   
Hacemos una primera inferencia estadística como vistazo general de las variables presentadas para el caso.

(A) Para estimar el salario de los ejecutivos (SALARY) en función de las ventas de la empresa (SALES), el precio de mercado (MKTVAL) y la antigüedad de la persona en el puesto (CEOTEN) se propone el siguiente modelo:

Y = B0 + B1X1 + B2X2 + B3X3 + U

Sin embargo, rápidamente detectamos que se están empleando variables con unidades de medida distintas, por lo que se propone primero estandarizar todas las variables y luego regresarlas de la misma manera. zVar serán las variables "estandarizadas"

Ejemplo: salary será zsalary

egen float zsales = std(sales), mean(0) std(1)

egen float zsalary = std(salary), mean(0) std(1)

egen float zmktval = std(mktval), mean(0) std(1)

egen float zceoten = std(ceoten), mean(0) std(1)

egen float zprofits = std(profits), mean(0) std(1)

reg zsalary zsales zmktval zceoten

Del modelo correctamente propuesto (con variables estandarizadas) llegamos a una conclusión distinta a la anterior:

Coeficientes:

- Por cada unidad que se incrementa sales, manteniendo el resto constante, se incrementa 0.2 la variable salario.

- Por cada unidad que se incrementa el valor de mercado de la empresa, ceteris paribus, se incrementa 0.25 la variable salario.

- Por cada unidad que se incrementa la antigüedad del CEO, manteniendo el resto de variables constantes, el salario se incrementa 0.15 unidades. Siendo esta la variable que menos afecta la explicación del salario del CEO.

Significatividad:

Significatividad:

- Para un pvalor de 0.06, la variable de ventas es significativa individualmente en el modelo.

- La variable MKTVAL es también significativa individualmente dentro del modelo, para un pvalor de 0.015 en el test t.

- Sin embargo, NO encontramos significancia individual en la variable de antigüedad, que para un pvalor de 0.025 no rechaza la hipótesis nula. Esto se puede constatar de la siguiente manera:

test zceoten=0 Haciendo un test F individual, podemos ver que NO rechaza la hipótesis para dicho pvalor

Significancia global:

- El modelo no es significativo en forma conjunta, debido a que el test F, para un pvalor cercano a cero, nos arroja un resultado de NO rechazo de hipótesis nula.

- Asimismo, el R2, es decir, el porcentaje de ajuste global del modelo, es muy bajo. Nos presenta una proporción de ajuste de 20%. Si regresáramos el modelo sin estandarizar, obtendríamos los mismos resultados de la salida de regresión

\*(B) Añadiremos la variable utilidad de la empresa (PROFITS) al modelo anterior.

reg zsalary zsales zmktval zceoten zprofits

De este nuevo modelo podemos extraer las siguientes conclusiones a partir de los resultados y comparaciones con el anterior:

Coeficientes:

- Mediante la adhesión de la variable utilidades, los coeficientes B0, B1, B2, B3 se mantuvieron casi inalterados en sus valores nominales.

Significatividad:

- De manera similar a la variable de antigüedad, las utilidades de la empresa no rechazan la hipótesis nula del test t, para un pvalor de 0.86, por lo que no son significativas individualmente.

- El test F de significancia global también NO rechaza la hipótesis nula, lo que hace no significativo al modelo en conjunto.

- El R2 es similar y si comparamos los R2 ajustados por los grados de libertad, tenemos que son modelos similares, con una diferencia infinitesimal estadísticamente de que el primer modelo es mejor (0,1828<0,1874).

\*(C) Evaluaremos presencia de multicolinealidad para los casos presentados:

corr zsales zmktval zceoten zprofits

De la matriz de correlación de variables observamos los siguientes resultados:

Valor de mercado-Ventas: correlación alta de 0.75

Utilidades-Ventas: correlación alta de 0.80

Valor de mercado-Utilidades: correlación alta de 0.92, casi perfecta.

Antigüedad-Utilidades: correlación negativa, imperfecta

Antigüedad-Ventas: correlación negativa, imperfecta

Valor de mercado-Antigüedad: correlación casi negativa

estat vif

Determinando los factores inflacionarios de varianza podemos ver que ningún VIF>10 por lo que no son valores conflictivos.

La media de FIV es 4.47, que de todas maneras sigue siendo un valor elevado (pero no problemático).

Por relación matemática, como ningún FIV > 10; también sabemos que ningún índice de tolerancia es menor a 0.1 (1/VIF<0.1)

vce, corr

Si observamos la matriz de correlación de los COEFICIENTES de la regresión, vemos que:

para aquellas variables que tomaban una correlación positiva, sus coeficientes están correlacionados negativamente.

Para las variables que tenían una correlación negativa, sus coeficientes adoptan una corr positiva.

Esta tendencia es un argumento que sustenta más nuestra hipótesis de multicolinealidad entre variables, aunque no es un determinante.

En síntesis, como no estamos en presencia de micro numerosidad (muestra pequeña) y los principales indicios concretos de multicolinealidad, como por ejemplo un R2 muy alto, alguna variable con VIF mayor a 10, etc. no podemos afirmar con seguridad que haya multicolinealidad, al menos NO perfecta.