**TP9 - Ejercicio 2: Heterocedasticidad conocida (MCP)**

La variable “fallas” representa la cantidad de fallas en la producción y “h\_uso” es la variable que almacena los datos de horas de utilización de la máquina de producción.

**Punto (A):** Corremos los siguientes códigos para la resolución;

*reg fallas h\_uso  
predict u, r  
gen u2 = u^2  
gen h\_uso2 = h\_uso^2  
reg u2 fallas h\_uso  
scatter u2 h\_uso*

*\* scatter u2 h\_uso2 (No son necesarios, pero tenemos la posibilidad también de graficarlos)  
\* scatter u2 fallas*

Regresamos las fallas contra las horas de uso, de allí extramoes los residuales de la regresión, los elevamos al cuadrado y hacemos una nueva variable. Esta es la que utilizaremos para graficar los errores al cuadrado para cada valor que toma la variable explicativa Xi.

Entonces, cuando graficamos los residuos al cuadrado contra los valores de la variable explicativa X (las horas de uso) podemos observar que hay poca dispersión de los residuos en los valores de Xi al principio y luego aumenta cuando cambia Xi, lo que nos indica presencia de heterocedasticidad.

Dado que la varianza de los errores deja de ser constante para los distintos valores de X, para poder corregirla, una propuesta sería rehacer el modelo general por mínimos cuadrados ponderados (MCP) que nos permitirá hallar estimadores más eficientes y errores homoscedasticos.

Conociendo la forma que toman los residuos al cuadrado respecto X, podemos inferir cómo será el término que h(x) que utilizaremos para corregir la ecuación: utilizaremos la raíz cuadra del término H, que dividirá cada elemento de la regresión.

**Punto (B):** Corremos los siguientes códigos para la resolución;

*gen H = h\_uso2  
gen rH = H^(12)*

*gen const\_H = 1rH* Generamos la constante, que no es más que B0 multiplicando el cociente de la raíz de H  
*gen fallas\_H = fallasrH* Generamos la variable explicada corregida por el término H   
*gen h\_uso\_H = h\_usorH* Generamos la variable explicativa corregida por el término H

*reg fallas\_H const\_H h\_uso\_H, noconst*

Corremos la regresión nueva que contiene la ecuación del modelo original pero corregida por el término H

**Punto (C):**

De los estimadores obtenidos por MCP de la regresión anterior, donde tenemos el siguiente modelo general:   
Fallas = (115.95) + (0.042)X ; podemos inferir que:

- Los estimadores son, por propiedad estadística, mucho más eficientes que los del modelo original.  
- En relación a su insesgadez, son estimadores insesgados, ya que esa propiedad GM la cumplían en su modelo original y con la transformación, los supuestos de Gauss-Markov que ya se cumplían, siguen cumpliéndose.

Solo cambia que ahora también se cumple el supuesto de errores homocedasticos.

**Punto (D):**

Si bien tenemos un R^2 de 0.971 para esta regresión, raramente nos serviría para utilizar como medida de bondad de ajuste, ya que las variables transformadas no explican lo que literalmente estamos estudiando detrás.

Para este caso, deberíamos seguir utilizando como medida de ajuste el R^2 del modelo original.

Cabe destacar como recordatorio, que la presencia de heteroscedasticidad no afecta nuestro R^2, por lo que podemos seguir empleando el original, que es de 0.6472 y representa un ajuste de la X sobre Y de casi un 65%.

**Punto (E):**

Para el caso donde la función de heterocedasticidad es incorrecta, es decir, que la función de varianza esté mal especificada, hay dos consecuencias:

Primeramente, caemos en la problemática de que los errores estándar y los estadísticos de prueba ya no son válidos, inclusive en muestras grandes.

Por otro lado, la otra consecuencia es que no sabremos con ninguna certeza que los estimadors MCP sean más eficientes que los de MCO.

**Punto (F):** Asumiendo hipoteticamente que la funcion H(x) que hemos empleado en el punto (B) está mal especificada, podemos proceder convirtiendo los errores estándar en robustos de la siguiente manera:

*reg fallas\_H const\_H h\_uso\_H, noconst r  
di "Diferencia " 23.11787-23.14042   
di "Diferencia " .0082973-.0077361*

Cuando convertimos los errores en robustos, podemos evaluar el tipo de diferencia que presentan frente a los errores estándar del modelo inicial por MCP.

La diferencia de los errores de la constante es negativa pero cercana a cero, además, la diferencia de los errores estándar de la variable X (horas de uso) es prácticamente cero, por lo que no tiene significatividad estadística.