Problema sobre heterocedasticidad 1.

Un investigador en el campo de la Economía de la Salud dispone del número de pacientes atendidos, *N*, y del gasto anual incurrido, *G*, para 50 Centros de Salud españoles y estima por MCO el siguiente modelo:

$$\hat{G}_i = 16,486 + 1,027N_i - 1,394N_i^2$$

Sospechando la presencia de heterocedasticidad, ajusta dos regresiones: una con los datos correspondientes a los 19 Centros de Salud con menor número de pacientes atendidos y otra con los datos de los 19 Centros con un mayor número de éstos. Las Sumas de Cuadrados de los Residuos obtenidas son 7 y 54, respectivamente.

El investigador define una nueva variable, *GN*, gasto por paciente, resultando la estimación:

$$\hat{GN}_i = 1,076 - 1,187N_i + 15,213INVN_i$$

donde $INVN_i = {}^{1}\!/_{N_i}$. De forma análoga al caso anterior, ajusta nuevamente dos regresiones, obteniendo que SCR_1=400 y SCR_2=750.

a) Contraste la presencia de heterocedasticidad en ambos modelos ($\alpha = 0,05$).

b) Explique el motivo por el que el investigador plantea el segundo modelo.

Solución

a) El contraste de hipótesis en este caso es:

$$\begin{cases} H_0: \text{Homocedasticidad} \\ H_1: \text{Heterocedasticidad} \end{cases}$$

A partir de la información proporcionada en el enunciado del ejercicio, la detección de la heterocedasticidad se puede llevar a cabo mediante el test de Goldfeld-Quandt, para el cual el estadístico F toma los siguientes valores:

- Para el modelo 1: $F_{\exp} = \dfrac{SCR_2 / gl_2}{SCR_1 / gl_1} = \dfrac{54/16}{7/16} = 7,714$

- Para el modelo 2: $F_{\exp} = \dfrac{SCR_2 / gl_2}{SCR_1 / gl_1} = \dfrac{750/16}{400/16} = 1,875$

Como el valor teórico de $F_{16,16}$ es 2,33 para un nivel de significación del 5%, se tiene que:

- Para el modelo 1: Como $F_{\exp} = 7,714 > F_{teo} = 2,33 \Rightarrow$ Se rechaza la hipótesis nula de homocedasticidad.

- Para el modelo 2: Como $F_{\exp} = 1,875 < F_{teo} = 2,33 \Rightarrow$ No se puede rechazar la hipótesis nula de homocedasticidad.

b) Partiendo del modelo original, para el cual existe heterocedasticidad, se puede suponer que la varianza de las perturbaciones es proporcional a N_i^2, es decir, que $E\left(u_i^2\right) = \sigma^2 N_i^2$. Así, dividiendo entre N_i, el término de perturbación del nuevo modelo es homocedástico:

$$E\left(v_i^2\right) = E\left(\frac{u_i^2}{N_i^2}\right) = \frac{\sigma^2 N_i^2}{N_i^2} = \sigma^2$$