Práctica 1.2: Simulación y el Modelo de Regressión

1. Ayudar a los niños de tres años a adquirir capacidades para el aprendizaje tiene unos beneficios potenciales elevados y persistentes. Supón que una institución de investigación diseña un experimento social que consiste en proporcionar tres horas adicionales de instrucción a niños pequeños. El instituto aleatoriza la admisión al programa: de todos los solicitantes, solo se aceptan un 30 % de los niños que son elegidos aleatoriamente. Considera el siguiente proceso generador de datos para los progresos cognitivos de los niños pequeños:

$$grades = 6 + 0{,}07 * program + 0{,}25 * family + u$$

donde

$$u \sim uniform(0, 2)$$

$$family \sim uniform(0, 1)$$

$$program = \begin{cases} 1 \text{ with probability } 0{,}3 \\ 0 \text{ with probability } 0{,}7 \end{cases}$$

Fija la seed a 457 para la replicabilidad. Fija el tamaño de la población en 5000 observaciones. Fija el número de replicaciones en 500. Para cada replicación:

a) Genera, en este orden, $u, family, program, grades$.

b) Para muestras de 100 y 5000 observaciones, estima $cov(program, family)$ y $cov(program, u)$.

c) Para muestras de 100 y 5000 observaciones, obtén las estimaciones MCO del modelo

$$grade = \gamma_0 + \gamma_1 program + u$$

Describe la distribución de $\hat{\gamma}_1$ para cada tamaño muestral. Comenta los resultados a la vista del resultado obtenido en (b).

2. En este ejercicio vamos a simular los resíduos obtenidos con estimación MCO de un modelo de regresión simple. Supón que tenemos dudas acerca de si el ratio t de un coeficiente sigue una distribución normal para muestras pequeñas. Simulamos el modelo remuestreando sobre los residuos obtenidos en la estimación MCO con los datos reales. La técnica de simular a partir de muestras reales se conoce como "bootstrapping" (para una presentación interesante de contrastes basados en la simulación y el bootstrapping, puedes consultar el capítulo 4 de Davidson and MacKinnon 2004). En este ejercicio vamos a realizar un bootstrap sencillo añadiendo el modelo estimado a los resíduos simulados (bootstrap paramétrico). En `gretl` la función resample() implementa remuestreo con reemplazo de una variable: dada una variable `x`, el comando `genr xr = resample(x)` crea una nueva variable en la que sus observaciones han sido obtenidas extrayéndolas aleatoriamente y con reemplazo de la variable original. Para ejecutar bootstrapping, vamos a utilizar los datos del fichero foodchild.gdt, que corresponden a una muestra de hogares en un país en vías de desarrollo. Los economistas están interesados en la relación entre el gasto total en alimentos en los hogares y el número de niños en esos hogares, así que vamos a estimar por MCO con desviaciones típicas robustas a heteroscedasticidad un modelo en el que el gasto total de alimentos (medido en dólares) es función lineal del número de niños en el hogar:

$$food\_exp_i = \beta_0 + \beta_1 n\_child_i + u_i$$

a) Usando desviaciones típicas robustas a heteroscedasticidad reporta el ratio t de la pendiente y su p-valor bajo la hipótesis nula de que $\beta_1 = 0$. Comenta los resultados.

b) Usando los resultados MCO, simula 500 veces los resíduos del modelo. Usando estos resíduos y las estiamciones de los coeficientes, construye valores simulados para $food\_exp_i$ bajo la nula de que $\beta_1 = 0$ (es decir, añade la estimacion MCO de la constante $\hat{\beta}_0$ a los resíduos simulados). En cada simulación, obtén el estimador de la pendiente y el ratio t.

c) Computa la desviación típica de los 500 estimaciones de la pendiente y compara el valor con la estimación robusta del apartado (a).

d) ¿Cuál es la proporción de casos en que el ratio t es superior al obtenido en el apartado (a). Comenta el resultado.