

Tarea 1 - Econometría

Otoño 2022

Fecha Entrega: 2 de Mayo, 18:00hrs.

La siguiente tarea es de carácter grupal. Esto no significa que se la deben dividir y cada integrante enfocarse en una parte, sino que deben discutir como grupo todas las partes de la tarea, y llegar a una solución entre todos/as. Es por lo tanto fundamental que cada integrante del grupo entienda las respuestas de cada sección, ya que las tareas forman parte de la materia del curso.

Cada grupo deberá desarrollar un informe de la tarea y enviarlo, junto con el m.file correspondiente, por email tanto a la profesora como a los ayudantes del curso.

La fecha límite de entrega es el día 2 de Mayo hasta las 18:00hrs. El atraso en la entrega de las tareas tendrá una penalización en la nota.

1 Inesegamiento y Multicolinealidad

1. Simule una base de datos que sigue el proceso generador de datos descrito a continuación:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + U_i$$

Para simular esto, utilice la siguiente informacin:

$$\begin{aligned} X_{1i} &\sim \mathcal{N}(0, 1) \\ X_{2i} &= 2X_{1i} + \nu_i \quad \text{donde} \quad \nu_i \sim \mathcal{N}(0, \sigma_\nu^2) \\ X_{3i} &\sim \mathcal{N}(0, 1) \\ U_i &= \delta_0 + \delta_1 X_{2i} + \delta_2 X_{3i} + \varepsilon_i \quad \text{donde} \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2) \end{aligned}$$

Utilice para su simulación los siguientes parámetros:

$$\begin{aligned} \beta &= \begin{pmatrix} 1 \\ 2 \\ 4 \\ 6 \end{pmatrix} \\ \delta &= \begin{pmatrix} 0 \\ 5 \\ 7 \end{pmatrix} \\ N &= 1000 \\ \sigma_\nu^2 &= 1 \\ \sigma_\varepsilon^2 &= 1 \end{aligned}$$

2. Estime una regresión de MCO de la variable Y sobre las variables X_1, X_2 y X_3 . Obtenga los coeficientes de MCO y sus errores estándar.
3. Estime una regresión de MCO de la variable Y sobre la variable X_1 y X_2 . Obtenga los coeficientes de MCO y sus errores estándar. Explique las diferencias obtenidas con respecto a los estimadores en el punto anterior.
4. Estime una regresión de MCO de la variable Y sobre la variable X_1 . Obtenga los coeficientes de MCO y sus errores estándar. Explique las diferencias obtenidas con respecto a los estimadores en el punto anterior.
5. Ahora vuelva a simular sus datos, manteniendo los mismos valores para X_1, X_2 y X_3 , pero ahora usando $\delta = 0$ para simular el término de error. Repita las estimaciones de 2., 3. y 4. con los nuevos datos, y comente sobre diferencias encontradas entre las dos estimaciones.
6. Vuelva a estimar una regresión de la variable Y sobre las variables X_1, X_2 y X_3 . Utilice un test t para testear las siguientes hipótesis nulas: $H_0 : \beta_1 = 0$ y $H_0 : \beta_2 = 0$. ¿Son los parámetros estadísticamente significativos? Utilice un gráfico para mostrar como los tests varían a medida que cambia σ_ν^2 .
7. Utilice un test F para testear la siguiente hipótesis nula: $H_0 : \beta_1 = \beta_2 = 0$. ¿Son los parámetros estadísticamente significativos? Utilice un gráfico para mostrar como el test varía a medida que cambia σ_ν^2 .

2 Forma Funcional y Efectos Marginales

Para esta parte de la tarea usaremos los datos de datos de SIMCE 2019 de la Agencia de Calidad de la Educación y una base de datos con el porcentaje de estudiantes prioritarios/as¹ por establecimiento, construida a partir de la base de datos de Alumnos Prioritarios, Preferentes y/o Beneficiarios de la Subvención Escolar Preferencial (SEP) del MINEDUC.² Ambas bases de datos están a nivel de establecimiento. Puede descargar las bases de datos en los siguientes links: BaseSIMCE y BaseSEP (formato de columnas separadas por coma). También está disponible el libro de códigos de la base de SIMCE en el siguiente link: GlosasSIMCE.

El objetivo es estudiar cómo el porcentaje de estudiantes prioritarios/as tiene un impacto en el SIMCE promedio. Para lo anterior, se le pide lo siguiente:

1. Importe ambas bases de datos y únalas, de modo que a la base de SIMCE se le agregue la información sobre porcentaje de estudiantes prioritarios/as. Quédese con las observaciones para las que dispone de información tanto en la base del SIMCE como en la base de estudiantes prioritarios/as. Realice estadística descriptiva de las variables a utilizar.

¹El artículo 2 de la Ley 20.248, establece que se considera como “Prioritarios”, a los alumnos para quienes la situación socioeconómica de sus hogares dificulte sus posibilidades de enfrentar el proceso educativo.

²Las bases de datos originales las pueden encontrar en los siguientes links: <https://datosabiertos.mineduc.cl> y <https://informacionestadistica.agenciaeducacion.cl>.

2. Realice una regresión del puntaje SIMCE del establecimiento en el porcentaje de estudiantes prioritarios. Utilice los controles que estime relevantes. Obtenga los estimadores de MCO, sus errores estándar y el R^2 . Interprete sus resultados.
3. Ahora genere la variable $prioritarios = prom_prioritarios \times 100$. Estime una regresión ahora utilizando *prioritarios* en lugar de *prom_prioritarios*. Obtenga los estimadores de MCO, sus errores estándar y el R^2 . Interprete los coeficientes estimados. Explique las diferencias entre las dos regresiones.
4. Dado que el efecto del porcentaje de estudiantes prioritarios/as puede no ser constante, se propone agregar dicha variable al cuadrado como un regresor adicional. Obtenga los estimadores de MCO, sus errores estándar y el R^2 . Interprete los coeficientes estimados.
5. ¿Cuál es el efecto marginal del porcentaje de estudiantes prioritarios/as en el SIMCE? Grafique el efecto marginal.
6. ¿Es el efecto marginal del porcentaje de estudiantes prioritarios/as en el SIMCE estadísticamente significativo? Haga un intervalo de confianza para el efecto marginal y grafíquelo.

3 Errores Estándar

Para las preguntas que siguen, vamos a utilizar la base de datos *DDK2011*. Pueden descargar dicha base de datos en el siguiente link: <https://www.ssc.wisc.edu/~bhansen/econometrics/>. Esta base de datos fue utilizada por Esther Duflo, Pascaline Dupas, y Michael Kremer en su paper “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya” publicado en *American Economic Review* el año 2011 (es altamente recomendable leer el artículo). En este paper se estudia el efecto del “tracking” (donde se separa a los estudiantes a una sala de clases de acuerdo a sus notas). El año 2005, 140 escuela en Kenia fueron financiadas para contratar un profesor/a extra de primero básico para reducir el tamaño de la sala de clases. En la mitad de las escuelas, los estudiantes fueron asignados a la sala de clases de acuerdo a sus notas iniciales. En la otra mitad de las escuelas, los estudiantes fueron asignados de forma aleatoria a las salas de clases.

1. Haga una regresión del rendimiento académico en tracking. Calcule los siguientes errores estándar:
 - (a) Errores estándar robustos.
 - (b) Errores estándar agrupados por colegio.
2. Ahora haga una regresión del rendimiento académico en tracking, edad, género, asignación al profesor/a contratado/a, y el percentil del/de la estudiante en la evaluación inicial. Calcule nuevamente ambos tipos de errores estándar. Discuta cómo cambian sus resultados.
3. ¿Es posible estimar la regresión anterior agregando efectos fijos por colegio? Discuta sus resultados.
4. Suponga ahora que quiere saber si el efecto de tracking es mayor para niños o para niñas. ¿Cómo lo haría para testarlo?