



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS
ICS2563 - ECONOMETRÍA APLICADA
SECCIÓN 1

Tarea 1

14 de marzo de 2022

2022-1 - Profesor Patricio Domínguez

Aspectos generales

- La tarea puede ser desarrollada en forma individual o en parejas (2 estudiantes).
- La fecha de entrega de la tarea es el **24 de Marzo a las 19:59 hrs** en el portal del curso en Canvas.
- La entrega debe incluir el script de R (o el dofile de STATA) y un informe de análisis en formato pdf.
- Sobre el **informe**:
 - El informe debe incorporar las respuestas a todas las preguntas, incorporando las figuras y/o tablas que estime conveniente.
- Sobre el **código**:
 - Cada entrega debe incluir el script de R (o el dofile de STATA) desde donde se corre todo el análisis.
 - Cada análisis debe ser desarrollado por ustedes mismos utilizando vectores y matrices; el uso de paquetes estadísticos será penalizado. Se indicará en las preguntas cuando deben realizar cálculos “a mano”, es decir, utilizando fórmulas vistas en clases, vectores y matrices.
 - Es importante que el código esté adecuadamente comentado para facilitar su corrección. Por ejemplo, indicar qué parte del informe/preguntas se desarrollan en cada sección del código.
 - **Reproducibilidad:** El código debiera estar escrito de manera tal que cualquier persona lo pueda correr, y reproducir los resultados desde su computador. En el caso de crear datos aleatorios recomendamos fijar una semilla que permita reproducir los resultados tal cual queden especificados en el informe de reporte.

- Parte de la tarea es que se vean enfrentados a tomar decisiones. Si deben aplicar criterio, háganlo y justifiquen sus elecciones.
- Ante dudas o preguntas, se recomienda fuertemente la utilización del foro de preguntas del curso. Las respuestas pueden servir a otros compañeros, que se enfrentan a las mismas dudas. No se permite publicar respuestas.
- La tarea tiene un total de **80 puntos**, en donde la presentación del informe posee 5 puntos (El informe debe incorporar las respuestas a todas las preguntas, y las figuras y/o tablas que estime necesarias), y la presentación del script 5 puntos (todos los cálculos realizados).
- **Bonificación:** Como un incentivo al uso del procesador de texto LATEX, se entregarán 5 puntos adicionales a quienes lo usen para escribir su informe.

Descripción de la tarea y pasos previos:

El objetivo de esta tarea es familiarizarnos con la estimación MCO utilizando datos del Ministerio de Educación de Chile (MINEDUC). La idea central es que puedan estimar la correlación entre el nivel de *vulnerabilidad social* de los establecimientos escolares y el puntaje en una prueba estandarizada (SIMCE). Si bien no existe una medida objetiva de *vulnerabilidad social* a nivel de establecimiento educacional, utilizaremos una variable *proxy* por construir a partir de la proporción de estudiantes de bajos ingresos según la información disponible por el MINEDUC.

En la práctica, para esta tarea deben juntar (*merge*) distintas bases de datos (BDD) a nivel de colegios. En cada una de las BDD hay un identificador de establecimientos escolares (*RBD*) que le permitirán unir la información proveniente de diferentes fuentes. En concreto, debe conectar las siguientes BDD:

- 1) SIMCE 8 básico 2019: Puntaje promedio SIMCE de todos los establecimientos que rindieron la prueba ese año. Descargar desde este [Link](#).
- 2) Matrícula por establecimiento: Número de estudiantes por establecimiento reconocidos por el MINEDUC. Descargar desde este [Link](#). y poner atención a variable *mat_total*.
- 3) Datos de estudiantes prioritarios por establecimiento: Número de estudiantes prioritarios para recibir la Subvención Escolar Preferencial (SEP) por establecimiento. Descargar desde este [Link](#) y poner atención a variable *n_prio*.

A continuación, lea los diccionarios de las bases de datos para identificar las variables presentes en cada una de ellas. Luego, crea una nueva variable “proporción estudiantes

prioritarios” dividiendo la cantidad de estudiantes prioritarios por el total de estudiantes matriculados en cada colegio. Utilizaremos esta variable como una variable proxy del índice de *vulnerabilidad social* del establecimiento educacional. Finalmente, inspeccione la BDD y familiarícese con las variables y sus respectivas observaciones.

Preguntas

1. Descripción de los datos (5 puntos)

1. Muestre un histograma de colegios según la proporción de estudiantes prioritarios. ¿Qué “proporción estudiantes prioritarios” tiene un establecimiento ubicado en la mediana de la distribución?
2. Estandarice las variables de puntaje SIMCE-Matemáticas. En adelante, utilice los valores estandarizados obtenidos como variable puntaje-SIMCE.

2. Brechas y esperanza incondicional (5 puntos)

Calcule las brechas de puntaje SIMCE-matemáticas de los establecimientos particulares pagados en relación con los particular subvencionados y municipales.

3. Scatterplot *v.s* Binscatter (15 puntos)

1. Muestra visualmente la relación entre “proporción estudiantes prioritarios” y puntaje SIMCE-matemáticas utilizando un scatterplot.
2. Construya **manualmente** un binscatter de 20 observaciones.
3. Compare ambas figuras y discuta.

4. MCO (20 puntos)

1. Estime una regresión lineal **en forma manual** (indicando paso a paso en el código) entre “proporción estudiantes prioritarios” y el puntaje SIMCE-matemáticas utilizando la base de datos completa a nivel de colegios.
2. Estime también los coeficientes **en forma manual** (indicando paso a paso en el código) utilizando las 20 observaciones utilizadas en el binscatter de la pregunta 3.2.
3. Compare los resultados en 4.1 y 4.2 y discuta cómo interpreta el coeficiente de “proporción estudiantes prioritarios” en cada caso.

5. Análisis de dispersión (15 puntos)

1. Seleccione una muestra aleatoria de 100 establecimientos y vuelva a estimar el coeficiente MCO de “proporción estudiantes beneficiados” sobre SIMCE-matemáticas. Repita este proceso 10,000 veces y muestre un histograma de los valores obtenidos del coeficiente. ¿Qué porcentaje de los valores son menores que 0? ¿Qué porcentaje de los valores son menores que -2.4?
2. Seleccione una muestra aleatoria de 1000 establecimientos y vuelva a estimar el coeficiente “proporción estudiantes beneficiados” sobre SIMCE-matemáticas. Repita este proceso 10,000 veces y muestre un histograma de los valores obtenidos del coeficiente. ¿Qué porcentaje de los valores son menores que 0? ¿Qué porcentaje de los valores son menores que -2.4?
3. ¿Qué podría decir al comparar los resultados en 5.1 y 5.2?

6. Análisis de política social (10 puntos)

Imagine Ud. que se desarrolla una política de equidad de financiamiento de la educación por estudiante (*voucher*) que reconoce la dificultad adicional que enfrentan colegios que reciben una proporción mayor de estudiantes en contexto de vulnerabilidad social. Para ello, el Gobierno asigna una subvención preferencial de recursos a establecimientos subvencionados y municipales que permitirían mejorar la *performance* en SIMCE-matemáticas en una cantidad equivalente al impacto en aprendizaje (SIMCE) que genera aumentar la cantidad de estudiantes de contexto de vulnerabilidad social.

Utilice el análisis de regresión realizado y asuma que el coeficiente estimado representaría el efecto causal en puntaje SIMCE-matemáticas asociado a aceptar una proporción determinada de estudiantes “beneficiarios” o de sectores socioeconómicos vulnerables. Calcule, en este nuevo escenario cuáles serían las brechas de puntaje SIMCE-matemáticas de los establecimientos particulares pagados en relación con los particular subvencionados y municipales.