**Evaluación de Impacto**

**MECA 4402**

**Profesor:** Rachid Laajaj

**Taller 1**

Regresión e Interpretación Causal

En su artículo [**“The Power of Example: Corruption Spurs Corruption”**](https://www.dropbox.com/s/tprlq4ybo61uwzr/Ajzenman%20%282021%29.pdf?dl=0)[[1]](#footnote-1), Nicolás Ajzeman (2021) investiga la influencia de la corrupción de políticos sobre la proporción de estudiantes que hacen trampa en las pruebas nacionales de México. Para eso, el autor realiza distintas regresiones con y sin controles a partir de las cuales encuentra evidencia que indica que la corrupción política incrementa la incidencia de trampa por parte de los estudiantes. En este taller ustedes realizarán algunas de esas estimaciones e interpretarán sus resultados. Para esto, deben usar la base de datos ***corruption.dta***, la cual es una submuestra aleatoria de la base usada por el autor.[[2]](#footnote-2)

Se recomienda tener respuestas tan breves como sea posible. Ciertas preguntas, sobre todo aquellas que preguntan por su opinión, pueden no tener una única respuesta correcta. Así, lo importante es que argumenten de manera coherente.

1. La especificación principal del autor consiste en regresar la proporción de estudiantes que hicieron trampa en el colegio , en el grado , del municipio en el año (prop) contra:
   1. Una dicótoma que toma el valor de uno si hubo corrupción en el municipio en el año (Corrupt).
   2. Una dicótoma que toma el valor de uno si fueron publicados reportes de auditorías[[3]](#footnote-3) en el municipio en el año (Auditada).
   3. Efectos fijos de colegio (clavedelaescuela).
   4. Efectos fijos de tiempo (year).
   5. El grado de los estudiantes (grade).
   6. El partido político activo (PartidoDesf).
   7. Una dicótoma que toma el valor de uno si el municipio fue auditado en el pasado (AlreadyAudited).
   8. Una dicótoma que toma el valor de uno si el municipio fue corrupto en el pasado (CorruptPast).
   9. El número de homicidios per cápita a nivel de municipio-año (HOMI\_CAP\_MUN).
   10. El logaritmo del total de impuestos recolectados en el municipio el año (total). [[4]](#footnote-4)
   11. Una dicótoma que toma el valor de uno si el partido político del municipio en el año está alineado con el nacional (MismoPartidoG).

Escriban matemáticamente la regresión a estimar y expliquen claramente a qué corresponde cada término.[[5]](#footnote-5) Asegúrense de escribir claramente a qué nivel de agregación está cada una de las variables incluidas en la regresión.[[6]](#footnote-6) Así mismo, indiquen cuál es el parámetro de interés el cual responde la pregunta de investigación del autor. De ser preferible, pueden resumir las variables de control – y solo las variables de control[[7]](#footnote-7) – en un único vector, sin embargo, deben explicar qué contiene dicho vector y cuál es su nivel de agregación.

1. Corra las siguientes estimaciones y en una tabla presente los resultados de las siguientes regresiones[[8]](#footnote-8):
   1. Una regresión simple de la variable dependiente contra la independiente principal.
   2. Una regresión simple de la variable dependiente contra la independiente principal + Controles.
   3. Una regresión simple de la variable dependiente contra la independiente principal + Efectos fijos.
   4. Una regresión simple de la variable dependiente contra la independiente principal + Controles + Efectos fijos.

La tabla debe tener el formato de presentación tipo artículo y estar completamente en español. Asegúrense de que esta tabla presente únicamente el coeficiente asociado a la variable de interés – i.e., no presente los coeficientes asociados a los controles, el intercepto o los efectos fijos. Debe ser claro, sin embargo, qué es incluido en cada columna. Para esto, pueden seguir el [siguiente formato de presentación](https://www.dropbox.com/s/u7rr1ort5c55r3k/Tabla_Ejemplo.PNG?dl=0).

1. Interpreten los resultados encontrados a partir de la primera regresión (estimación punto 2, inciso b). ¿Parece haber una relación entre la corrupción política y la trampa de los estudiantes? ¿Podemos interpretar esta relación como causal? ¿Cuál sería el supuesto de identificación? Argumenten.
2. Escojan la especificación que – a su criterio – parece ser la más adecuada para aproximarse a la respuesta de la pregunta de investigación del autor y respondan:
   1. ¿Cuál es el supuesto de identificación para que esta especificación les permita hacer interpretaciones causales?
3. Dados los resultados de la especificación que escogieron en el anterior punto: interpreten el coeficiente estimado.

1. Esto es un enlace al artículo en cuestión. [↑](#footnote-ref-1)
2. La base de datos la pueden encontrar en Bloque Neón. [↑](#footnote-ref-2)
3. Según el autor, las auditorías fueron llevadas a cabo con el fin de investigar si en la contratación pública hubo ejecuciones o gastos no autorizados que pueden ser enmarcados como corrupción. Así, cuando se publica un reporte de auditoría se sabe si hubo o no corrupción en un contrato en el pasado. [↑](#footnote-ref-3)
4. La variable en la base de datos está en niveles. Ustedes deben crear el logaritmo de la variable. [↑](#footnote-ref-4)
5. Por ejemplo, si escriben la regresión a estimar de la siguiente forma:

   Entonces, deben escribir explícitamente a qué corresponde cada variable. Es decir, si la regresión es de talla para la edad contra ingesta calórica + controles + efectos fijos, deben escribir algo del estilo “ captura la talla para la edad del estudiante del colegio , corresponde a la ingesta calórica promedio del estudiante del colegio en el último año y es un vector de controles. Este último contiene [var1], [var2], … corresponde a efectos fijos de colegio y es el término del error”. [↑](#footnote-ref-5)
6. Por ejemplo, “la variable corresponde a la talla para la edad del estudiante del colegio en el periodo , mientras que corresponde a los ingresos promedio de los estudiantes del colegio en el periodo ”. [↑](#footnote-ref-6)
7. Esto quiere decir que la variable explicativa principal debe estar fuera de dicho vector, así como los efectos fijos. [↑](#footnote-ref-7)
8. En STATA pueden usar el comando regress y en R la función lm(). En STATA existen un comando alternativo que permite controlar de forma eficiente por efectos fijos: reghdfe. Deben usar la opción absorb() para incluir los efectos fijos. Por ejemplo, si quieren incluir efectos fijos de colegio, el comando debe ser reghdfe y x1 … xk, absorb(código\_colegio). Para exportar los resultados a Word, Excel o Latex, en STATA pueden usar los comandos outreg u outreg2. En R la función stargazer permite exportar los resultados a Latex. [↑](#footnote-ref-8)