



EVALUACIÓN DE IMPACTO

RACHID LAAJAJ

Taller 1- Regresión e Interpretación Causal

Repositorio: <https://github.com/zagarciab/T1_ADI>

Elaborado por:

Nicolas Jacome Velasco
Zaira Alejandra Garcia
Bernal

22 agosto de 2025

1. Especificación de la regresión

La especificación principal del autor consiste en regresar la proporción de estudiantes que hicieron trampa en el colegio c , en el grado g , del municipio m en el año t ($prop$) contra variables de control. De tal forma que, se puede realizar una aproximación matemática de la regresión que el Nicolás Ajzeman (2021) indica como:

$$Prop_{sgmt} = \beta_0 + \beta_1 Corrupt_{mt} + \beta_2 Auditada_{mt} + X_{sgmt}\Gamma + \lambda_s + \delta_t + \epsilon_{is} \quad (1)$$

Donde la variable dependiente:

- $Prop_{sgmt}$: captura la proporción de estudiantes que hicieron trampa. Está a nivel colegio (s), grado (g), municipio (m) y año (t). Es un variable numérico entre 0 y 1.

Parámetros y variables de interés

- $Corrupt_{mt}$: corresponde a la variable de interés dicotómica que indica uno si hubo corrupción. Su nivel de agregación es municipio (m) y año (t).
- $Auditada_{mt}$: corresponde a la variable dicotómica que indica si hubo reportes de auditoria. Su nivel de agregación es municipio (m) y año (t).
- β_1 y β_2 : son los parámetros de interés que responden a la pregunta de investigación, los cuales miden el efecto de la corrupción y las auditorías, respectivamente, sobre la proporción de estudiantes que hacen trampa.

Variables de control

$X_{sgmt}\Gamma$ corresponde a un vector de controles y sus coeficientes correspondientes. Este vector contiene:

- $Grado_{gt}$: es el grado de los estudiantes, a nivel grado (g) y año (t).
- $PartidoDesf_{mt}$: es el partido político activo, a nivel municipio (m) y año (t).
- $AlredyAudited_{mt}$: es binaria, que toma uno para las auditorías anteriores a la fecha, a nivel municipio (m) y año (t).
- $CorruptPast_{mt}$: es binaria que toma uno para corrupción en el pasado, a nivel municipio (m) y año (t).
- $HOMICAPMUN_{mt}$: son homicidios per cápita, a nivel municipio (m) y año (t).

Efectos fijos

- λ_s : corresponde a efectos fijos a nivel de colegio s . Controla por todas las características no observadas que son constantes en el tiempo dentro de cada colegio (clavedelaescuela).
- δ_t : corresponde a los efectos fijos del tiempo (year). Controla por todas las características no observadas que son comunes a todos los colegios en un año determinado a nivel de año (t).

Términos adicionales

- β_0 : es la constante del modelo.

- ϵ_{is} : corresponde al término del error que captura todas las variables no observadas que afectan a la proporción de estudiantes que hacen trampa.

2. Resultados de estimación de regresiones

En el repositorio de GitHub o en el script se podrá validar la exploración de la base de datos realizada, en donde se identificó missing values para las variables Prop y Unauthorized se imputaron los missing values con la mediana. Por otra parte, se imputaron missing values con la moda para las variables: Auditada, AlreadyAudited, CorruptPast, Corrupt, HOMI_CAP_MUN y GradoSecundaria.

a. Una regresión simple de la variable dependiente contra la independiente principal

$$Prop_{sgmt} = \beta_0 + \beta_1 Corrupt_{mt} + \epsilon_{is} \quad (2)$$

b. Una regresión simple de la variable dependiente contra la independiente principal + controles

$$Prop_{sgmt} = \beta_0 + \beta_1 Corrupt_{mt} + \beta_2 Auditada_{mt} + X_{sgmt}\Gamma + \epsilon_{is} \quad (3)$$

c. Una regresión simple de la variable dependiente contra la independiente principal + efectos fijos

$$Prop_{sgmt} = \beta_0 + \beta_1 Corrupt_{mt} + \beta_2 Auditada_{mt} + \lambda_s + \delta_t + \epsilon_{is} \quad (4)$$

d. Una regresión simple de la variable dependiente contra la independiente principal + controles + efectos fijos

$$Prop_{sgmt} = \beta_0 + \beta_1 Corrupt_{mt} + \beta_2 Auditada_{mt} + X_{sgmt}\Gamma + \lambda_s + \delta_t + \epsilon_{is} \quad (5)$$

Tabla de resultados de las estimaciones de los puntos a, b, c, d.

Cuadro 1: Resultados de las regresiones

	Dependent variable:			
	OLS		panel linear	
	MCO	MCO con Controles	E.F.	E.F con Controles
	(1)	(2)	(3)	(4)
Corrupción	0.001 (0.001)	0.0005 (0.001)	0.004*** (0.001)	0.003** (0.002)
CorruptPast1		0.002* (0.001)		-0.004** (0.002)
Controles	No	Sí	No	Sí
Efectos fijos	No	No	Sí	Sí
Observations	102,640	102,637	102,640	102,640
R ²	0.00001	0.033	0.0002	0.014
Adjusted R ²	0.00000	0.032	-0.250	-0.234
Residual Std. Error	0.125 (df = 102638)	0.123 (df = 102546)		
F Statistic	1.351 (df = 1; 102638)	38.388*** (df = 90; 102546)	14.127*** (df = 1; 82116)	12.780*** (df = 88; 82029)

Note:

*p<0.1; **p<0.05; ***p<0.01
Errores estándar entre paréntesis. Solo se reporta el coeficiente de 'Corrupt'.

3. Interpretación de resultados

Interpretar 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?

El coeficiente estimado para la variable Corrupción política es positivo pero no resulta estadísticamente significativo. Esto indica que, una vez controladas las demás variables (características de la escuela, auditorías, contexto municipal, etc.), no se encuentra evidencia robusta de una asociación entre corrupción política y la proporción de estudiantes que hacen trampa. Aunque el coeficiente tiene signo positivo, su magnitud es muy pequeña (aprox. 0.00046), adicionalmente, su p-valor es de 0.737284, el cual es superior a los umbrales de significancia de 0.05. Lo anterior sugiere que, aun de existir un efecto, este sería prácticamente irrelevante en términos estadísticos y sustantivos debido a que contradice las expectativas de que una cultura de deshonestidad a nivel político se refleje como trampa de los estudiantes.

¿Podemos interpretar esta relación como causal?

Este resultado no puede interpretarse como evidencia causal. El hecho de que la corrupción no sea significativa en esta especificación puede deberse a problemas de medición, variables omitidas o correlaciones no observadas que distorsionan la estimación. Incluso si hubiera significancia, la estimación seguiría siendo solo una correlación.

Lo anterior, se basa al considerar principalmente el problema de variables omitidas, ya que para una relación causal, sería importante considerar variables constantes en el tiempo dentro de la escuela como objeto de estudio, tales como: cultura estudiantil y lineamientos de comportamiento establecidos por las escuelas, entre otros; asimismo, controlar por factores comunes en todas las escuelas que hacen parte del estudio.

Adicionalmente, el modelo demuestra endogeneidad al materializarse la causalidad inversa debido a la variable Auditada1. Para este caso, el coeficiente de la variable es de 0.004618 y presenta significancia estadística, lo cual indica que las auditorías pueden ser enfocadas en instituciones con sospecha de trampa, y esa misma sospecha estar correlacionada con la trampa estudiantil.

¿Cuál sería el supuesto de identificación?

El supuesto de identificación es que el valor esperado de la trampa que habrían tenido los colegios con corrupción si no hubieran tenido corrupción es igual al valor esperado de la trampa observada en los colegios sin corrupción. En otras palabras, se asume que, en ausencia de corrupción, los colegios corruptos se comportarían igual que los no corruptos. Bajo ese supuesto, las diferencias observadas en trampa pueden atribuirse únicamente a la corrupción. Sin embargo, el siguiente supuesto es bastante fuerte: es probable que los colegios corruptos y no corruptos difieran en otras características (recursos, nivel socioeconómico, ubicación, etc.), y esas diferencias también afectan la trampa. Por eso la identificación causal es débil en este caso.

4. Mejor especificación

A continuación, se elige la regresión de Efectos Fijos con Controles (4) en la tabla resumen, la cual parece ser la más adecuada para aproximarse a la respuesta de la pregunta de investigación del autor. Esta especificación combina controles y efectos fijos, lo que ayuda a mitigar la endogeneidad. Incluye controles observables, lo que reduce el riesgo de omitir factores que puedan estar correlacionados con corrupción y con el desempeño de los estudiantes. Además, incorpora efectos fijos de municipio y de año, lo que permite controlar por factores no observados que son constantes en el tiempo y por choques comunes a todos en un año. De esta manera, se reduce el sesgo por variables omitidas y se aproxima mejor a una interpretación causal.

Adicionalmente, para dar respuesta al supuesto de identificación para que esta especificación permita hacer interpretaciones causales es posible considerar que el supuesto de identificación es que, dados los controles y los efectos fijos, la corrupción es independiente del error; es decir, los colegios expuestos a corrupción sirven como buen contrafactual de sí mismos en ausencia de corrupción. En términos de resultados potenciales: el valor esperado del resultado potencial de trampa en un colegio si no hubiera corrupción es igual al valor esperado del resultado observado en colegios sin corrupción, condicionado a los controles y efectos fijos.

El supuesto de identificación es:

$$E[Y_i(0)|D_i = 1, X, FE] = E[Y_i(0)|D_i = 0, X, FE] \quad (6)$$

5. Interpretación de coeficiente estimado

Teniendo en cuenta la especificación anterior (punto 4) se interpreta que el coeficiente estimado para la variable `Corrupt1` en el modelo de efectos fijos con controles (Modelo 4) es de 0.0030483. Este valor es estadísticamente significativo a un nivel de confianza del 5

Esto significa que, manteniendo constantes los efectos fijos de la escuela y del año, y controlando por las demás variables, el hecho de que un municipio pase a ser considerado corrupto.^{est} está asociado con un aumento promedio de 0.003 puntos porcentuales en la proporción de estudiantes que hacen trampa. Este es un efecto positivo, lo que indica que la corrupción municipal se relaciona con un mayor comportamiento de trampa en las escuelas, pero la magnitud del efecto es muy pequeña en términos prácticos. La significancia estadística sugiere que esta relación no es producto del azar y que las variables de omitidas estaban ocultando esta relación en los modelos más simples.