

Taller 1 - Regresión e Interpretación Causal

Nicolas Jacome Velasco, Zaira Alejandra Garcia Bernal

2025-08-16

Evaluación de Impacto

Taller 1: Regresión e Interpretación Causal

0. Carga de datos y paquetes

```
# Instalar
if (!requireNamespace("pacman", quietly = TRUE)) {
  install.packages("pacman")
}
```

```
# Cargar los paquetes necesarios
pacman::p_load(haven, plm, skimr, dplyr, ggplot2)
```

```
library(haven)
library(plm)
library(skimr)
library(dplyr)
library(haven)
library(ggplot2)
```

```
# Carga de datos .dta
url_data <- "https://github.com/zagarciab/T1_ADI/raw/main/3.%20Data/corruption_SV.dta"

base <- read_dta(url_data)
```

0.1 Exploración de los datos

```
# Estructura de datos y estadísticas
str(base)
```

```
## tibble [102,640 x 16] (S3: tbl_df/tbl/data.frame)
## $ clavedelaescuela: chr [1:102640] "01DES00010" "01PES0037H" "01DTV0095Z" "01PES0004Q" ...
## ..- attr(*, "label")= chr "Unique Key per School"
## ..- attr(*, "format.stata")= chr "%11s"
## $ turno           : num [1:102640] 1 1 1 1 2 1 1 1 2 2 ...
## ..- attr(*, "label")= chr "morning/evening/night"
```

```
##   .-. attr(*, "format.stata")= chr "%10.0g"
##   $ prop           : num [1:102640] 0.00893 0 0 0 0.00901 ...
##   .-. attr(*, "label")= chr "Proportion of Students Cheating"
##   .-. attr(*, "format.stata")= chr "%9.0g"
##   $ GradoSecundaria : num [1:102640] 3 3 3 3 3 3 3 3 3 ...
##   .-. attr(*, "label")= chr "First, Second or Third Year of Secondary School"
##   .-. attr(*, "format.stata")= chr "%9.0g"
##   $ year           : num [1:102640] 2006 2006 2006 2006 2006 ...
##   .-. attr(*, "format.stata")= chr "%9.0g"
##   $ clave_mun       : chr [1:102640] "01001" "01001" "01001" "01001" ...
##   .-. attr(*, "format.stata")= chr "%78s"
##   $ unauthorized    : num [1:102640] 0 0 0 0 0 0 0 0 0 ...
##   .-. attr(*, "label")= chr "% of unauthorized expenditure"
##   .-. attr(*, "format.stata")= chr "%10.0g"
##   $ Auditada        : num [1:102640] 0 0 0 0 0 0 0 0 0 ...
##   .-. attr(*, "format.stata")= chr "%9.0g"
##   $ AlreadyAudited  : num [1:102640] 0 0 0 0 0 0 0 0 0 ...
##   .-. attr(*, "format.stata")= chr "%9.0g"
##   $ CorruptPast     : num [1:102640] 0 0 0 0 0 0 0 0 0 ...
##   .-. attr(*, "format.stata")= chr "%9.0g"
##   $ Corrupt         : num [1:102640] 0 0 0 0 0 0 0 0 0 ...
##   .-. attr(*, "label")= chr "1 if unauthorized>0, 0 otherwise"
##   .-. attr(*, "format.stata")= chr "%9.0g"
##   $ HOMI_CAP_MUN    : num [1:102640] 2.5e-05 2.5e-05 2.5e-05 2.5e-05 2.5e-05 ...
##   .-. attr(*, "format.stata")= chr "%9.0g"
##   $ total           : num [1:102640] 1.74e+09 1.74e+09 1.74e+09 1.74e+09 1.74e+09 ...
##   .-. attr(*, "format.stata")= chr "%10.0g"
##   $ mis_tot         : num [1:102640] 0 0 0 0 0 0 0 0 0 ...
##   .-. attr(*, "format.stata")= chr "%9.0g"
##   $ PartidoDesf     : chr [1:102640] "PAN" "PAN" "PAN" "PAN" ...
##   .-. attr(*, "label")= chr "Partido"
##   .-. attr(*, "format.stata")= chr "%30s"
##   $ MismoPartidoG   : num [1:102640] 1 1 1 1 1 1 1 1 1 ...
##   .-. attr(*, "format.stata")= chr "%9.0g"
```

```
skimr::skim_with(numeric = list(
  p25 = ~round(quantile(., 0.25, na.rm = TRUE), 2),
  p50 = ~round(quantile(., 0.5, na.rm = TRUE), 2),
  p75 = ~round(quantile(., 0.75, na.rm = TRUE), 2),
  mean = ~round(mean(., na.rm = TRUE), 2),
  sd = ~round(sd(., na.rm = TRUE), 2)
))
```

```
## Creating new skimming functions for the following classes: p25, p50, p75, mean, sd.
## They did not have recognized defaults. Call get_default_skimmers() for more information.
```

```
## function (data, ..., .data_name = NULL)
## {
##   if (is.null(.data_name)) {
##     .data_name <- rlang::expr_label(substitute(data))
##   }
##   if (!inherits(data, "data.frame")) {
##     data <- as.data.frame(data)
```

```

##   }
##   stopifnot(inherits(data, "data.frame"))
##   selected <- names(tidyselect::eval_select(rlang::expr(c(...)),
##     data))
##   if (length(selected) == 0) {
##     selected <- names(data)
##   }
##   grps <- dplyr::groups(data)
##   if (length(grps) > 0) {
##     group_variables <- selected %in% as.character(grps)
##     selected <- selected[!group_variables]
##   }
##   else {
##     attr(data, "groups") <- list()
##   }
##   skimmers <- purrr::map(selected, get_final_skimmers, data,
##     local_skimmers, append)
##   types <- purrr::map_chr(skimmers, "skim_type")
##   unique_skimmers <- reduce_skimmers(skimmers, types)
##   combined_skimmers <- purrr::map(unique_skimmers, join_with_base,
##     base)
##   ready_to_skim <- tibble::tibble(skim_type = unique(types),
##     skimmers = purrr::map(combined_skimmers, mangle_names,
##       names(base$funcs)), skim_variable = split(selected,
##         types)[unique(types)])
##   grouped <- dplyr::group_by(ready_to_skim, .data$skim_type)
##   nested <- dplyr::summarize(grouped, skimmed = purrr::map2(.data$skimmers,
##     .data$skim_variable, skim_by_type, data))
##   structure(tidy::unnest(nested, "skimmed"), class = c("skim_df",
##     "tbl_df", "tbl", "data.frame"), data_rows = nrow(data),
##     data_cols = ncol(data), df_name = .data_name, dt_key = get_dt_key(data),
##     groups = dplyr::group_vars(data), base_skimmers = names(base$funcs),
##     skimmers_used = get_skimmers_used(unique_skimmers))
## }
## <bytecode: 0x000001afffc6f3c0>
## <environment: 0x000001afffc61820>

```

```
skim(base)
```

```

## Warning: There was 1 warning in `dplyr::summarize()`.
## i In argument: `dplyr::across(tidyselect::any_of(variable_names),
##   mangled_skimmers$funcs)`.
## i In group 0: .
## Caused by warning:
## ! There were 4 warnings in `dplyr::summarize()`.
## The first warning was:
## i In argument: `dplyr::across(tidyselect::any_of(variable_names),
##   mangled_skimmers$funcs)`.
## Caused by warning in `grepl()`:
## ! unable to translate '12KTV4<d0>905' to a wide string
## i Run `dplyr::last_dplyr_warnings()` to see the 3 remaining warnings.

```

Table 1: Data summary

Name	base
Number of rows	102640
Number of columns	16
Column type frequency:	
character	3
numeric	13
Group variables	None

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
clavedelaescuela	0	1	0	10	3	20516	0
clave_mun	0	1	0	5	3	576	0
PartidoDesf	0	1	0	11	3	83	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	hist
turno	3	1	1.140000e+00	0.000000e+00	1	1	1	1.000000e+00	4	
prop	507	1	4.000000e-02	1.300000e-01	0	0	0	2.000000e-02	1	
GradoSecundaria	3	1	2.150000e+00	0.000000e+00	1	1	2	3.000000e+00	3	
year	0	1	2.010440e+03	0.000000e+00	2006	2009	2011	2.012000e+03	2013	
unauthorized	3	1	2.220000e+00	0.000000e+00	0	0	0	0.000000e+00	100	
Auditada	3	1	3.200000e-01	4.700000e-01	0	0	0	1.000000e+00	1	
AlreadyAudited	3	1	7.100000e-01	4.500000e-01	0	0	1	1.000000e+00	1	
CorruptPast	3	1	4.700000e-01	5.000000e-01	0	0	0	1.000000e+00	1	
Corrupt	3	1	1.500000e-01	3.600000e-01	0	0	0	0.000000e+00	1	
HOMI_CAP_MUN	3	1	0.000000e+00	0.000000e+00	0	0	0	0.000000e+00	0	
total	3	1	1.004345e+09	0.009867e+09	0	145976853	375823221	1.487667e+09	7961572707	
mis_tot	3	1	4.000000e-02	2.000000e-01	0	0	0	0.000000e+00	1	
MismoPartidoG	0	1	3.200000e-01	4.700000e-01	0	0	0	1.000000e+00	1	

```
# Imputación de missing values
```

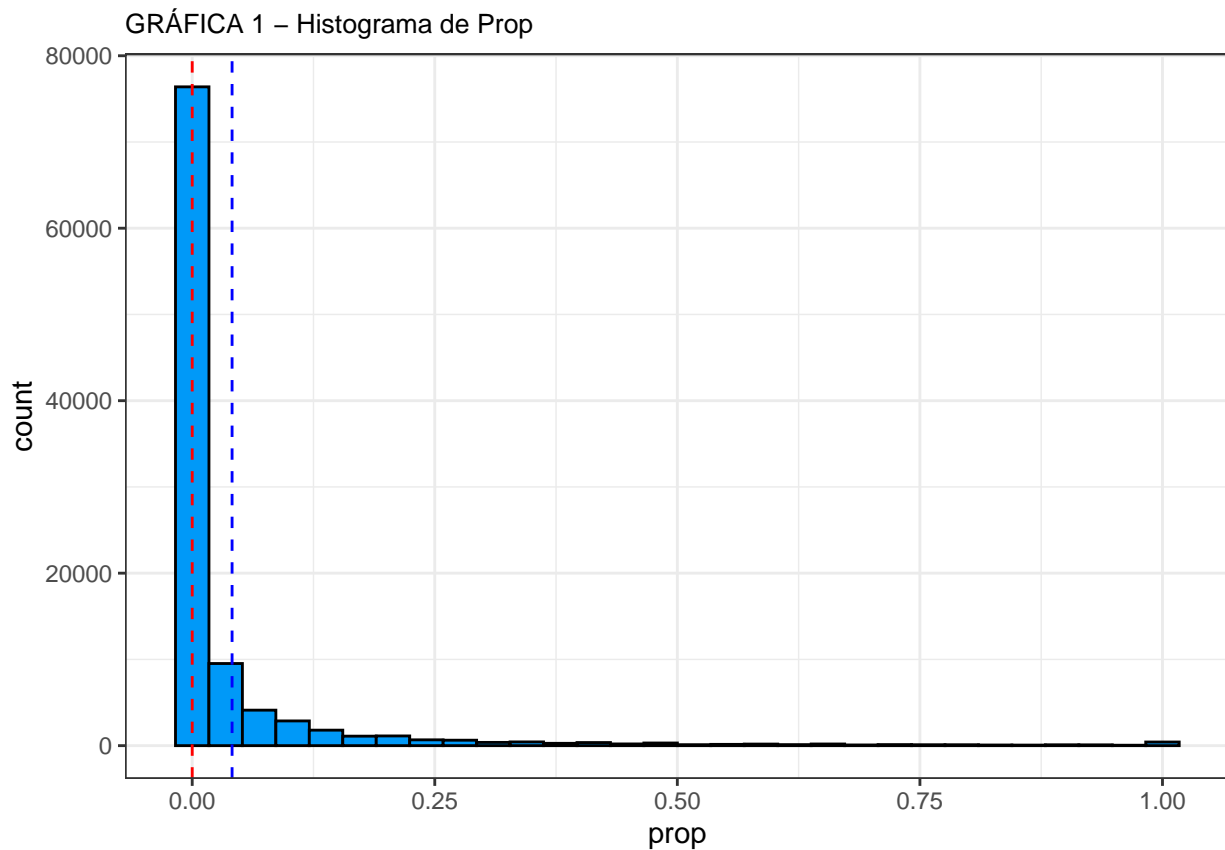
```
## Histogramas
```

```
ggplot(base, aes(prop)) +  
  geom_histogram(color = "#000000", fill = "#0099F8") +
```

```
geom_vline(xintercept = median(base$prop, na.rm = TRUE), linetype = "dashed", color = "red") +
geom_vline(xintercept = mean(base$prop, na.rm = TRUE), linetype = "dashed", color = "blue") +
ggtitle("GRÁFICA 1 - Histograma de Prop ") +
theme_bw() +
theme(plot.title = element_text(size = 10))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 507 rows containing non-finite outside the scale range
## (`stat_bin()`).
```

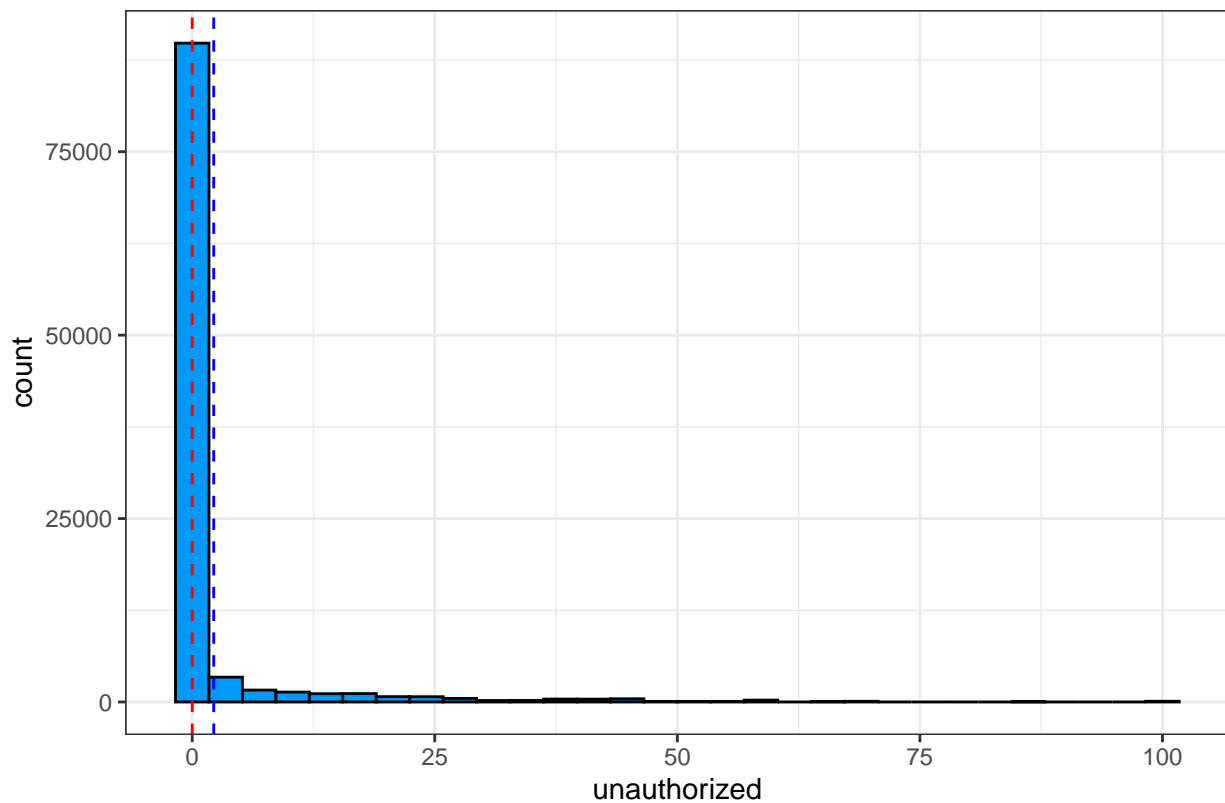


```
ggplot(base, aes(unauthorized)) +
  geom_histogram(color = "#000000", fill = "#0099F8") +
  geom_vline(xintercept = median(base$unauthorized, na.rm = TRUE), linetype = "dashed", color = "red") +
  geom_vline(xintercept = mean(base$unauthorized, na.rm = TRUE), linetype = "dashed", color = "blue") +
  ggtitle("GRÁFICA 2 - Histograma de unauthorized ") +
  theme_bw() +
  theme(plot.title = element_text(size = 10))
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

```
## Warning: Removed 3 rows containing non-finite outside the scale range
## (`stat_bin()`).
```

GRÁFICA 2 – Histograma de unauthorized



```
## Imputación de missing values con la mediana para Prop
base <- base %>%
  mutate(prop = ifelse(is.na(prop) == TRUE,
                        median(base$prop, na.rm = TRUE) , prop),
         unauthorized = ifelse(is.na(unauthorized) == TRUE,
                               median(base$unauthorized, na.rm = TRUE),
                               unauthorized))

## Imputación variables dicotómica con moda
moda_auditada <- base %>%
  filter(!is.na(Auditada)) %>%
  count(Auditada) %>%
  arrange(desc(n)) %>%
  slice(1) %>%
  pull(Auditada)

moda_AlreadyAudited <- base %>%
  filter(!is.na(AlreadyAudited)) %>%
  count(AlreadyAudited) %>%
  arrange(desc(n)) %>%
  slice(1) %>%
  pull(AlreadyAudited)

moda_CorruptPast <- base %>%
  filter(!is.na(CorruptPast)) %>%
  count(CorruptPast) %>%
```

```

    arrange(desc(n)) %>%
    slice(1) %>%
    pull(CorruptPast)

moda_Corrupt <- base %>%
  filter(!is.na(Corrupt)) %>%
  count(Corrupt) %>%
  arrange(desc(n)) %>%
  slice(1) %>%
  pull(Corrupt)

moda_HOMI <- base %>%
  filter(!is.na(HOMI_CAP_MUN)) %>%
  count(HOMI_CAP_MUN) %>%
  arrange(desc(n)) %>%
  slice(1) %>%
  pull(HOMI_CAP_MUN)

moda_grado <- base %>%
  filter(!is.na(GradoSecundaria)) %>%
  count(GradoSecundaria) %>%
  arrange(desc(n)) %>%
  slice(1) %>%
  pull(GradoSecundaria)

moda_grado <- base %>%
  filter(!is.na(GradoSecundaria)) %>%
  count(GradoSecundaria) %>%
  arrange(desc(n)) %>%
  slice(1) %>%
  pull(GradoSecundaria)

base <- base %>%
  mutate(Auditada = ifelse(is.na(Auditada) == TRUE,
                           moda_auditada, Auditada),
         AlreadyAudited = ifelse(is.na(AlreadyAudited) == TRUE,
                                   moda_AlreadyAudited, AlreadyAudited),
         CorruptPast = ifelse(is.na(CorruptPast) == TRUE,
                               moda_CorruptPast, CorruptPast),
         Corrupt = ifelse(is.na(Corrupt) == TRUE,
                           moda_Corrupt, Corrupt),
         HOMI_CAP_MUN = ifelse(is.na(HOMI_CAP_MUN) == TRUE,
                                moda_HOMI, HOMI_CAP_MUN),
         GradoSecundaria = ifelse(is.na(GradoSecundaria) == TRUE,
                                    moda_grado, GradoSecundaria))

## Convertir a factor variables categoricas
var_factor <- c("clavedelaescuela", "turno", "GradoSecundaria", "year",
               "clave_mun", "Auditada", "AlreadyAudited", "CorruptPast",
               "Corrupt")

base <- base %>%
  mutate(across(all_of(var_factor), as.factor))

```

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:

$$Prop_{sgmt} = \beta_0 + \beta_1 Corrupt_{mt} + \beta_2 Auditada_{mt} + X_{sgmt}\Gamma + \lambda_s + \delta_t + \varepsilon_{is}$$

Donde la variable dependiente:

- $Prop_{sgmt}$ caputura 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.

Donde los parámetros y variables de interés son:

- $Corrupt_{mt}$ 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 ($grado$) 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).

Donde, los efectos fijos son:

- 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

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

```
lm1<- lm (prop ~ Corrupt, base)
summary(lm1)
```

```
##
## Call:
## lm(formula = prop ~ Corrupt, data = base)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.04194 -0.04068 -0.04068 -0.02308  0.95932
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.0406767  0.0004252  95.664  <2e-16 ***
## Corrupt1     0.0012603  0.0010844   1.162    0.245
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1253 on 102638 degrees of freedom
## Multiple R-squared:  1.316e-05, Adjusted R-squared:  3.418e-06
## F-statistic: 1.351 on 1 and 102638 DF, p-value: 0.2451
```

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

```
# Regresión por MCO con controles
lm2 <- lm (prop ~ Corrupt + Auditada + GradoSecundaria + PartidoDesf + AlreadyAudited + CorruptPast + HOMI_CAP_MUN + total + MismoPartidoG,
summary(lm2)
```

```
##
## Call:
## lm(formula = prop ~ Corrupt + Auditada + GradoSecundaria + PartidoDesf +
##      AlreadyAudited + CorruptPast + HOMI_CAP_MUN + total + MismoPartidoG,
##      data = base)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.16546 -0.04501 -0.02960 -0.00716  1.00393
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.259e-02  1.158e-02   1.952 0.050999 .
## Corrupt1          4.658e-04  1.389e-03   0.335 0.737284
## Auditada1         4.618e-03  1.158e-03   3.988 6.65e-05 ***
## GradoSecundaria2  1.775e-02  1.020e-03  17.412 < 2e-16 ***
## GradoSecundaria3  2.057e-02  9.474e-04  21.711 < 2e-16 ***
## PartidoDesfACDC   -3.102e-02  1.660e-02  -1.869 0.061680 .
## PartidoDesfALIANZA 8.760e-02  1.216e-02   7.204 5.90e-13 ***
## PartidoDesfAPA     8.064e-03  1.233e-02   0.654 0.513289
```

## PartidoDesfAPAG	6.679e-03	1.227e-02	0.544	0.586342	
## PartidoDesfAPM	-2.137e-02	1.214e-02	-1.761	0.078310	.
## PartidoDesfAPPJ	2.062e-03	1.633e-02	0.126	0.899534	
## PartidoDesfAPPT	-1.092e-02	1.385e-02	-0.788	0.430492	
## PartidoDesfAPT	-2.817e-02	1.323e-02	-2.128	0.033309	*
## PartidoDesfAPVM	-2.802e-02	2.246e-02	-1.248	0.212152	
## PartidoDesfASXXI	-1.550e-02	1.738e-02	-0.892	0.372416	
## PartidoDesfATT	7.827e-03	1.988e-02	0.394	0.693852	
## PartidoDesfAXB	-6.288e-03	1.211e-02	-0.519	0.603429	
## PartidoDesfAXCH	1.223e-02	1.278e-02	0.957	0.338720	
## PartidoDesfAXJ	-4.906e-03	1.192e-02	-0.412	0.680624	
## PartidoDesfAXM	-3.811e-03	1.298e-02	-0.294	0.769048	
## PartidoDesfAXTB	-2.064e-02	1.217e-02	-1.696	0.089827	.
## PartidoDesfAXU	9.410e-02	1.540e-02	6.110	1.00e-09	***
## PartidoDesfC.C.	-7.230e-03	1.159e-02	-0.624	0.532863	
## PartidoDesfCAC	-1.600e-02	1.635e-02	-0.979	0.327751	
## PartidoDesfCAFV	-1.504e-02	1.183e-02	-1.272	0.203488	
## PartidoDesfCCC	-2.963e-02	1.469e-02	-2.017	0.043702	*
## PartidoDesfCCCH	-1.381e-02	1.276e-02	-1.082	0.279382	
## PartidoDesfCCPQ	-1.946e-02	1.371e-02	-1.420	0.155618	
## PartidoDesfCCS	1.507e-04	1.219e-02	0.012	0.990133	
## PartidoDesfCLT	-6.679e-04	1.496e-02	-0.045	0.964391	
## PartidoDesfCOALICION	-8.942e-04	1.157e-02	-0.077	0.938377	
## PartidoDesfCONVER.	1.853e-02	1.274e-02	1.455	0.145766	
## PartidoDesfCPBT	5.104e-02	1.295e-02	3.941	8.12e-05	***
## PartidoDesfCPEM	-1.922e-02	1.191e-02	-1.614	0.106475	
## PartidoDesfCPG	4.749e-02	1.680e-02	2.826	0.004710	**
## PartidoDesfCPJ	-6.263e-03	1.258e-02	-0.498	0.618722	
## PartidoDesfCPM	-2.101e-02	2.768e-02	-0.759	0.447712	
## PartidoDesfCPNL	-2.106e-02	1.354e-02	-1.556	0.119733	
## PartidoDesfCPP	8.099e-03	1.189e-02	0.681	0.495730	
## PartidoDesfCPSL	-1.030e-02	1.814e-02	-0.568	0.569977	
## PartidoDesfCQRA	1.482e-02	1.554e-02	0.953	0.340349	
## PartidoDesfCSA	2.491e-02	1.205e-02	2.067	0.038744	*
## PartidoDesfCUPG	1.466e-02	1.197e-02	1.225	0.220752	
## PartidoDesfCUV	-3.507e-02	2.288e-02	-1.533	0.125364	
## PartidoDesfCUXT	-1.093e-03	1.210e-02	-0.090	0.928012	
## PartidoDesfCXBDC	4.785e-02	1.888e-02	2.534	0.011272	*
## PartidoDesfCXBDCH	4.488e-02	1.490e-02	3.012	0.002600	**
## PartidoDesfCXBDT	2.100e-02	2.024e-02	1.038	0.299439	
## PartidoDesfCXBNCH	3.125e-02	2.498e-02	1.251	0.210941	
## PartidoDesfCXBNLT	5.972e-02	1.841e-02	3.243	0.001181	**
## PartidoDesfCXBP	-6.637e-03	2.929e-02	-0.227	0.820704	
## PartidoDesfCXBT	3.731e-02	1.436e-02	2.599	0.009357	**
## PartidoDesfECV	6.710e-02	2.412e-02	2.782	0.005399	**
## PartidoDesfGNU	6.399e-02	1.408e-02	4.546	5.48e-06	***
## PartidoDesfJPC	-2.157e-02	1.543e-02	-1.398	0.162085	
## PartidoDesfJXBDT	-4.859e-03	1.372e-02	-0.354	0.723246	
## PartidoDesfJXNL	-8.236e-03	1.217e-02	-0.677	0.498463	
## PartidoDesfMATCQROO	6.668e-03	1.290e-02	0.517	0.605143	
## PartidoDesfMC	3.592e-02	2.601e-02	1.381	0.167268	
## PartidoDesfMNU	1.211e-01	3.742e-02	3.236	0.001211	**
## PartidoDesfMP	4.196e-02	2.410e-02	1.741	0.081714	.
## PartidoDesfMXH	-1.793e-02	1.245e-02	-1.440	0.150011	

```
## PartidoDesfNCCH      1.082e-02  1.589e-02   0.681 0.495956
## PartidoDesfNNU      -6.903e-03  1.289e-02  -0.536 0.592280
## PartidoDesfNS       -2.688e-02  1.245e-02  -2.159 0.030854 *
## PartidoDesfPAN       1.873e-03  1.150e-02   0.163 0.870618
## PartidoDesfPAS       1.590e-02  1.945e-02   0.818 0.413512
## PartidoDesfPNA       3.194e-03  1.373e-02   0.233 0.816021
## PartidoDesfPOR CHIAPAS 4.703e-02  6.273e-02   0.750 0.453469
## PartidoDesfPRD       1.804e-02  1.164e-02   1.551 0.121005
## PartidoDesfPRI       1.649e-02  1.154e-02   1.429 0.153101
## PartidoDesfPRIZAC    -3.569e-03  1.210e-02  -0.295 0.767958
## PartidoDesfPS        1.483e-02  2.722e-02   0.545 0.585879
## PartidoDesfPSD       1.164e-01  1.926e-02   6.042 1.52e-09 ***
## PartidoDesfPT        3.218e-02  1.210e-02   2.660 0.007812 **
## PartidoDesfPVEM       4.320e-02  1.197e-02   3.609 0.000307 ***
## PartidoDesfQROOAV.   -1.555e-03  1.758e-02  -0.088 0.929502
## PartidoDesfTEP       -3.454e-02  7.213e-02  -0.479 0.632041
## PartidoDesfUPI       -1.256e-02  2.721e-02  -0.462 0.644339
## PartidoDesfUPT       -8.985e-03  1.333e-02  -0.674 0.500284
## PartidoDesfUXC       6.038e-02  1.294e-02   4.666 3.07e-06 ***
## PartidoDesfUXCH      9.104e-02  1.225e-02   7.434 1.06e-13 ***
## PartidoDesfUYC      -1.378e-02  1.841e-02  -0.749 0.453999
## PartidoDesfXENQTQ    -1.866e-02  1.233e-02  -1.513 0.130293
## PartidoDesfXGR       -8.915e-03  1.198e-02  -0.744 0.456797
## PartidoDesfZANUNE     2.978e-02  1.367e-02   2.179 0.029335 *
## AlreadyAudited1     -1.590e-03  1.220e-03  -1.303 0.192621
## CorruptPast1         1.728e-03  1.041e-03   1.660 0.096993 .
## HOMI_CAP_MUN         5.578e+00  1.372e+00   4.065 4.81e-05 ***
## total               -4.862e-12  3.686e-13 -13.191 < 2e-16 ***
## MismoPartidoG        6.160e-05  1.271e-03   0.048 0.961337
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1233 on 102546 degrees of freedom
## (3 observations deleted due to missingness)
## Multiple R-squared:  0.03259, Adjusted R-squared:  0.03174
## F-statistic: 38.39 on 90 and 102546 DF, p-value: < 2.2e-16
```

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

```
# Matriz con efectos fijos
```

```
m_base <- pdata.frame(base, index = c("clavedelaescuela", "year"))
```

```
## Warning in pdata.frame(base, index = c("clavedelaescuela", "year")): duplicate couples (id-time) in :
## to find out which, use, e.g., table(index(your_pdataframe), useNA = "ifany")
```

```
# Modelo de efectos fijos (within estimator)
```

```
m_ef1 <- plm(prop ~ Corrupt,
```

```
data = m_base,
```

```
model = "within", effect = "twoways") # Estimación de modelo de EF y se especifica q
```

```
summary(m_ef1)
```

```
## Twoways effects Within Model
```

```
##
## Call:
## plm(formula = prop ~ Corrupt, data = m_base, effect = "twoways",
##      model = "within")
##
## Unbalanced Panel: n = 20516, T = 1-22, N = 102640
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -0.6614039 -0.0217740 -0.0021992  0.0085460  0.8900557
##
## Coefficients:
##              Estimate Std. Error t-value Pr(>|t|)
## Corrupt1 0.0043728   0.0011634   3.7586 0.000171 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1013.3
## Residual Sum of Squares: 1013.1
## R-Squared:    0.00017201
## Adj. R-Squared: -0.24971
## F-statistic: 14.1271 on 1 and 82116 DF, p-value: 0.00017098
```

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

```
m_ef2 <- plm(prop ~ Corrupt + Auditada + GradoSecundaria + PartidoDesf + AlreadyAudited + CorruptPast +
             data = m_base,
             model = "within", effect = "twoways")

summary(m_ef2)
```

```
## Twoways effects Within Model
##
## Call:
## plm(formula = prop ~ Corrupt + Auditada + GradoSecundaria + PartidoDesf +
##      AlreadyAudited + CorruptPast + HOMI_CAP_MUN, data = m_base,
##      effect = "twoways", model = "within")
##
## Unbalanced Panel: n = 20516, T = 1-22, N = 102640
##
## Residuals:
##      Min.      1st Qu.      Median      3rd Qu.      Max.
## -0.6688389 -0.0240442 -0.0039739  0.0123376  0.9007144
##
## Coefficients: (1 dropped because of singularities)
##              Estimate Std. Error t-value Pr(>|t|)
## Corrupt1          3.0483e-03  1.5528e-03  1.9631 0.0496432 *
## Auditada1          2.7161e-05  1.2346e-03  0.0220 0.9824490
## GradoSecundaria2    1.7188e-02  9.9173e-04 17.3313 < 2.2e-16 ***
## GradoSecundaria3    2.5546e-02  9.9552e-04 25.6611 < 2.2e-16 ***
## PartidoDesfAAA       1.8197e-02  1.8055e-02  1.0078 0.3135404
## PartidoDesfACDC     -2.0784e-03  1.9153e-02 -0.1085 0.9135874
```

## PartidoDesfALIANZA	2.6684e-02	1.1937e-02	2.2354	0.0253926	*
## PartidoDesfAPA	9.2004e-04	1.2707e-02	0.0724	0.9422809	
## PartidoDesfAPAG	1.5353e-02	1.2726e-02	1.2065	0.2276398	
## PartidoDesfAPM	2.9726e-03	1.1651e-02	0.2551	0.7986131	
## PartidoDesfAPPJ	1.8554e-02	1.8172e-02	1.0210	0.3072577	
## PartidoDesfAPPT	1.7688e-02	1.9621e-02	0.9015	0.3673352	
## PartidoDesfAPT	2.0035e-02	1.2687e-02	1.5792	0.1143053	
## PartidoDesfAPVM	-3.5133e-02	2.7301e-02	-1.2869	0.1981393	
## PartidoDesfASXXI	3.1121e-03	2.6405e-02	0.1179	0.9061783	
## PartidoDesfATT	4.1881e-02	2.4032e-02	1.7427	0.0813821	.
## PartidoDesfAXB	-3.3196e-02	1.3306e-02	-2.4948	0.0126056	*
## PartidoDesfAXCH	1.3864e-02	1.3060e-02	1.0616	0.2884308	
## PartidoDesfAXJ	-1.3179e-02	1.1572e-02	-1.1389	0.2547553	
## PartidoDesfAXM	1.4060e-02	1.5092e-02	0.9317	0.3515170	
## PartidoDesfAXTB	-8.1737e-03	1.2099e-02	-0.6756	0.4993040	
## PartidoDesfAXU	7.1300e-02	1.9145e-02	3.7242	0.0001961	***
## PartidoDesfC.C.	4.2249e-03	1.0886e-02	0.3881	0.6979508	
## PartidoDesfCAC	3.0314e-02	1.6546e-02	1.8321	0.0669344	.
## PartidoDesfCAFV	-8.8085e-03	1.1433e-02	-0.7705	0.4410262	
## PartidoDesfCCC	8.4114e-03	1.5105e-02	0.5569	0.5776125	
## PartidoDesfCCCH	2.7126e-03	1.3461e-02	0.2015	0.8403005	
## PartidoDesfCCPQ	9.2548e-03	1.3574e-02	0.6818	0.4953649	
## PartidoDesfCCS	2.8737e-03	1.2574e-02	0.2285	0.8192296	
## PartidoDesfCLT	1.4090e-02	1.8851e-02	0.7475	0.4547910	
## PartidoDesfCOALICION	-3.3833e-04	1.0856e-02	-0.0312	0.9751383	
## PartidoDesfCONVER.	1.5051e-02	1.2995e-02	1.1582	0.2467734	
## PartidoDesfCPBT	-3.1360e-02	1.2642e-02	-2.4807	0.0131160	*
## PartidoDesfCPEM	5.9261e-03	1.1325e-02	0.5233	0.6007912	
## PartidoDesfCPG	4.6472e-02	1.7846e-02	2.6040	0.0092153	**
## PartidoDesfCPJ	-2.2094e-03	1.2259e-02	-0.1802	0.8569804	
## PartidoDesfCPM	-9.1649e-04	3.1033e-02	-0.0295	0.9764399	
## PartidoDesfCPNL	-4.4530e-03	1.5706e-02	-0.2835	0.7767818	
## PartidoDesfCPP	2.0510e-03	1.2199e-02	0.1681	0.8664861	
## PartidoDesfCPSL	1.8698e-02	1.8406e-02	1.0158	0.3097039	
## PartidoDesfCQRA	-2.3889e-02	1.8794e-02	-1.2711	0.2037023	
## PartidoDesfCSA	2.3416e-02	1.2179e-02	1.9227	0.0545243	.
## PartidoDesfCUPG	1.1315e-02	1.2261e-02	0.9228	0.3561050	
## PartidoDesfCUV	-1.7094e-02	2.2790e-02	-0.7501	0.4532152	
## PartidoDesfCUXT	1.2259e-02	1.1806e-02	1.0383	0.2991105	
## PartidoDesfCXBDC	1.4841e-02	2.3029e-02	0.6445	0.5192762	
## PartidoDesfCXBDCH	1.1592e-02	1.5908e-02	0.7286	0.4662209	
## PartidoDesfCXBDT	-4.2983e-02	2.2316e-02	-1.9261	0.0540978	.
## PartidoDesfCXBNCH	-4.1936e-02	2.9635e-02	-1.4151	0.1570477	
## PartidoDesfCXBNLT	5.6590e-02	2.2723e-02	2.4904	0.0127611	*
## PartidoDesfCXBP	-1.3530e-02	3.5749e-02	-0.3785	0.7050882	
## PartidoDesfCXBT	2.0047e-02	1.4770e-02	1.3572	0.1747076	
## PartidoDesfECV	5.9766e-02	2.4906e-02	2.3997	0.0164129	*
## PartidoDesfGNU	4.0534e-02	1.3972e-02	2.9011	0.0037195	**
## PartidoDesfJPC	-6.3964e-03	1.8844e-02	-0.3394	0.7342765	
## PartidoDesfJXBDT	1.0887e-02	1.5695e-02	0.6937	0.4878750	
## PartidoDesfJXNL	2.1583e-03	1.4448e-02	0.1494	0.8812470	
## PartidoDesfMATCQROO	-1.4175e-02	1.4370e-02	-0.9865	0.3239039	
## PartidoDesfMC	3.0754e-02	2.7111e-02	1.1344	0.2566316	
## PartidoDesfMNU	1.5994e-02	4.1930e-02	0.3814	0.7028715	

```

## PartidoDesfMP          4.9801e-02  2.4981e-02  1.9936  0.0462013 *
## PartidoDesfMXH          3.5361e-03  1.2524e-02  0.2823  0.7776883
## PartidoDesfNCCH          1.3866e-02  1.9209e-02  0.7219  0.4703849
## PartidoDesfNNU          4.7816e-03  1.3566e-02  0.3525  0.7244861
## PartidoDesfNS           3.4100e-02  2.9141e-02  1.1702  0.2419264
## PartidoDesfPAN          9.7464e-03  1.0790e-02  0.9033  0.3663907
## PartidoDesfPAS          2.3921e-02  2.8768e-02  0.8315  0.4056825
## PartidoDesfPNA          2.0335e-02  1.3677e-02  1.4867  0.1370851
## PartidoDesfPOR CHIAPAS -4.4705e-02  6.4808e-02 -0.6898  0.4903212
## PartidoDesfPRD          6.0308e-03  1.0715e-02  0.5628  0.5735603
## PartidoDesfPRI          8.8661e-03  1.0685e-02  0.8298  0.4066784
## PartidoDesfPRIZAC       -1.4452e-03  1.1832e-02 -0.1221  0.9027880
## PartidoDesfPS           -1.1450e-02  3.1054e-02 -0.3687  0.7123404
## PartidoDesfPSD          1.1477e-01  2.2179e-02  5.1746  2.289e-07 ***
## PartidoDesfPT           4.3488e-02  1.3010e-02  3.3426  0.0008302 ***
## PartidoDesfPVM          1.6286e-02  1.1574e-02  1.4071  0.1594025
## PartidoDesfQROOAV.      8.2487e-04  2.1594e-02  0.0382  0.9695300
## PartidoDesfTEP          -3.9832e-03  1.2009e-01 -0.0332  0.9735399
## PartidoDesfUPI          -3.5944e-03  3.8031e-02 -0.0945  0.9247011
## PartidoDesfUPT          5.5919e-03  2.0234e-02  0.2764  0.7822676
## PartidoDesfUXC          3.5177e-02  1.3503e-02  2.6051  0.0091859 **
## PartidoDesfUXCH         4.2236e-02  1.1936e-02  3.5385  0.0004026 ***
## PartidoDesfUYC          -3.2988e-01  1.0135e-01 -3.2548  0.0011352 **
## PartidoDesfXENQTQ       5.6464e-03  1.2501e-02  0.4517  0.6515044
## PartidoDesfXGR          -4.4778e-02  1.3114e-02 -3.4144  0.0006395 ***
## AlreadyAudited1        -8.7867e-04  1.9712e-03 -0.4458  0.6557735
## CorruptPast1           -4.0269e-03  1.7845e-03 -2.2566  0.0240369 *
## HOMI_CAP_MUN           -4.6460e+00  2.4178e+00 -1.9216  0.0546649 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    1013.3
## Residual Sum of Squares: 999.61
## R-Squared:              0.013525
## Adj. R-Squared:        -0.23433
## F-statistic: 12.78 on 88 and 82029 DF, p-value: < 2.22e-16

```

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

```
library(stargazer)
```

```
##
```

```
## Please cite as:
```

```
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer
```

```

stargazer(lm1, lm2, m_ef1, m_ef2,
  type = "text",          # Cambiar a "html" o "latex" si quieres
  keep = "Corrupt",       # Solo muestra el coeficiente de interés
  column.labels = c("MCO", "MCO con Controles", "E.F.", "E.F con Controles"),

```

```

dep.var.labels = "Proporción",
covariate.labels = "Corrupción",
digits = 3,
no.space = TRUE,
title = "Resultados de las regresiones",
add.lines = list(
  c("Controles", "No", "Sí", "No", "Sí"),
  c("Efectos fijos", "No", "No", "Sí", "Sí")
),
notes = "Errores estándar entre paréntesis. Solo se reporta el coeficiente de 'Corrupt'."

```

```

##
## Resultados de las regresiones
## =====
##                                     Dependent variable:
## -----
##                                     Proporción
##                                     OLS                                     panel
##                                     linear
##                                     MCO      MCO con Controles      E.F.
##                                     (1)      (2)      (3)
## -----
## Corrupción      0.001      0.0005      0.004***
##                  (0.001)      (0.001)      (0.001)
## CorruptPast1      0.002*
##                  (0.001)
## -----
## Controles      No      Sí      No
## Efectos fijos      No      No      Sí
## Observations      102,640      102,637      102,640
## R2      0.00001      0.033      0.0002
## Adjusted R2      0.00000      0.032      -0.250
## 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) 12.1
## =====
## Note:
##                                     *p<0
##                                     Errores estándar entre paréntesis. Solo se reporta el co

```

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: $E[Y_i(0)|D_i = 1, X, FE] = E[Y_i(0)|D_i = 0, X, FE]$

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%, con un p-valor de 0.0496432.

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” 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.