

# Extensiones a OLS

Abril 17, 2024

---

Prof. Sergio Béjar

Departamento de Estudios Políticos, CIDE

# Plan para Hoy

*Vamos a hacer extensiones al modelo OLS que ya conocemos (i.e. regresión lineal bivariada).*

# Limitaciones de la Regresión Bivariada (o Simple)

La regresión bivariada nunca es suficiente.

- En ciencia política y relaciones internacionales hay muchas variables de interés que no son de intervalo (i.e. medidas en escala continua, ordenadas y con intervalos iguales entre sus valores).
- Regresión Bivariada no nos permite controlar por “confounders” (i.e. explicaciones alternativas).

Esta presentación nos ayudará a entender estos temas con más claridad.

# Variables Dicotómicas (Dummies)

Las variables dicotómicas (dummy) están por todos lados en ciencias políticas y RI.

- Juegan un papel importante en la regresión de efectos “fijos”.
- Algunas veces simplemente estamos interesados en el efecto de “una sola cosa”.

# Estados Pivotales (Swing States) y Turnout

Vamos a regresar al ejemplo de turnout. Pero ahora estamos interesados en entender el efecto en turnout de un estado que es “pivotal”. Variable “ss” en la base de datos.

- Asumamos (basados en los datos de 538) que los estados pivotales son: CO, FL, IA, MI, MN, NV, NH, NC, OH, PA, VA, y WI.
- Cuando  $x = 0$ , tenemos y-intercept.

```
M1 <- lm(turnoutho ~ ss, data=election_turnout)

library(stargazer)
stargazer(M1, style="ajps",
  omit.stat=c("F","rsq","ser"), header=FALSE,
  dep.var.labels.include = FALSE,
  covariate.labels=c("Estado Pivotal"),
  title="Efecto de Estado Pivotal en Turnout, 2016")
```

**Table 1:** Efecto de ser Estado Pivotal en Turnout, 2016

Estado Pivotal	7.371*** (1.747)
Constant	59.087*** (0.847)
N	51
Adj. R-squared	0.252

\*\*\*p < .01; \*\*p < .05; \*p < .1

# Estados Pivotales (Swing States) y Turnout

- El turnout estimado en estados “seguros” es 59.09%
- El turnout estimado en estados “pivotales” es 66.46%
- El efecto de un estado pivotal en turnout es **7.37%** (e.s. = 1.75)
- El valor de  $t$  es 4.22

Podemos inferir con alto grado de confianza que ser un estado pivotal tiene un efecto positivo en turnout.



# Variación Regional y Turnout

Los estados del Sur tienden a tener niveles de turnout bajos.

- La mayoría son estados “seguros”.
- Tienen alto porcentaje de pobreza que incrementa el costo de salir a votar.
- Tienen historia de amplias restricciones en el derecho a votar.

Vamos primero a desempacar “variación regional” viendo el efecto que tiene el Sur (relativo a no-Sur) en turnout.

**Table 2:** Efecto de ser un Estado del Sur en Turnout, 2016

Sur	−3.465* (1.768)
Constant	61.976*** (1.020)
N	51
Adj. R-squared	0.054

\*\*\* $p < .01$ ; \*\* $p < .05$ ; \* $p < .1$

## Variación Regional y Turnout

- El turnout estimado en estados que no son del sur es 61.98%.
- El turnout estimado en estados del sur es '58.51%.
- El efecto estimado del “Sur” es **-3.46%** .
- $t$ : -1.96

Los resultados sugieren en efecto negativo. - Pero no es un efecto muy grande. La diferencia es de aproximadamente 3%.

# Efectos Fijos de Región y Turnout

El modelo anterior no es muy informativo.

- Además, es un tanto problemático porque trata a todos los estados que no están en el sur como homogéneos.

Obviously, this last regression isn't that informative.

- It also problematically treats non-Southern states as homogenous.
- Una  $R^2$  baja sugiere que:

Podemos especificar otras regiones como “efectos fijos”.

- Estos modelos tratan a las variables explicativas como una serie de variables dummy para cada valor de  $x$ .
- Un grupo se deja afuera como “categoría base”
  - O no tendríamos intercepto en  $y$ .

**Table 3:** Efecto de Regiones en Turnout, 2016

Noreste	6.099** (2.351)
Medio-Oeste	4.805** (2.151)
Oeste	0.404 (2.102)
Constant	58.512*** (1.383)
N	51
Adj. R-squared	0.131

\*\*\*p < .01; \*\*p < .05; \*p < .1

# Efectos Fijos de Región y Turnout

Interpretación de esta regresión:

Todos los coeficientes indican el efecto de esa región **contra** la categoría base.

- La categoría base en este modelo es el Sur.
- El turnout estimado en el Sur es 58.51%.
- El turnout en el Noreste es sustancialmente más alto que en el Sur ( $t = 2.59$ ).
- El turnout en el Medio-Oeste es más alto que en el sur ( $t = 2.23$ ).
- No hay diferencia significativa importante entre el turnout en el Oeste y en el Sur ( $t = 0.19$ ).

# Regresión Múltiple

El ejemplo anterior (i.e. efectos fijos por región) es un ejemplo de **regresión múltiple**.

- Pero no tiene variables de control.

La regresión múltiple produce **coeficientes parciales de regresión**.

# Regresión Múltiple

Regresemos al ejemplo de turnout. Asumamos:

- $x_1$ : % de ciudadanos en el estado con licenciatura.
- $x_2$ : estados en el Sur.
- $x_3$ : estado es pivotal.

Importante: al añadir variables estamos “controlando” por confounders potenciales.



## Racional al Usar Variables de Control

Asumamos que uno de ustedes esta proponiendo un argumento en el que el nivel de educación a nivel estatal explica turnout. Yo podría decir que hay un **“sesgo por variables omitidas”** por lo siguiente:

- Ser un estado en el Sur disminuye educación y turnout.
- El efecto de ser un “estado pivotal” puede explicar el nivel de educación e incrementar turnout.

En otras palabras, yo sostengo que tu argumento relacionando educación ( $x$ ) con turnout ( $y$ ) es espurio a estos factores ( $z$ ).

- Para eso “controlamos”. No queremos explicar variación sino ver el efecto de los confounders.

**Table 4:** Comparación de Modelos

	<b>Model 1</b>	<b>Model 2</b>	<b>Model 3</b>
Universidad			0.384*** (0.111)
Estado Pivotal	7.371*** (1.747)		7.008*** (1.546)
Sur		-3.465* (1.768)	-1.940 (1.415)
Constant	59.087*** (0.847)	61.976*** (1.020)	48.479*** (3.468)
N	51	51	51
R-squared	0.266	0.073	0.455
Adj. R-squared	0.252	0.054	0.420

# Regresión Múltiple

- Turnout estimado para (1) estado no en el Sur, (2) no pivotal, y (3) nadie se gradúa de la universidad es 48.48%.
  - Parece razonable pero el valor mínimo de turnout en los datos es 19.2% (West Virginia).
- El coeficiente parcial de regresión para la variable “universidad” es .38 ( $t = 3.47$ ).
- El coeficiente parcial de regresión para la variable “Sur” es insignificante.
- El efecto estimado de ser un “estado pivotal” es de un incremento de  $\sim 7\%$  ( $t = 4.53$ ) en turnout.

# Efectos Interactivos

La regresión múltiple es lineal y aditiva.

- Sin embargo, algunos efectos (digamos:  $x_1$ ) pueden depender en el valor de otra variable (digamos:  $x_2$ ).

A esto le llamamos un **efecto interactivo**.

# Ejemplo del Mundo Real

Consideremos el argumento de Zaller (1992):

- Democrats son ligeramente más pro-aborto que los Republicanos.
- Esa diferencia es condicional en que tanto los individuos entienden de política.

Usaremos datos de ANES 2012 para evaluar esta hipótesis.

# Nuestros Datos

**Variables Independientes:** Partidismo (Party ID), conocimiento político, interacción entre las dos variables.

- Partidismo: (0 = Dem, 1 = Independiente, 2 = Rep)
- Conocimiento político: la persona sabe quién es el Orador de la Cámara de Representantes?

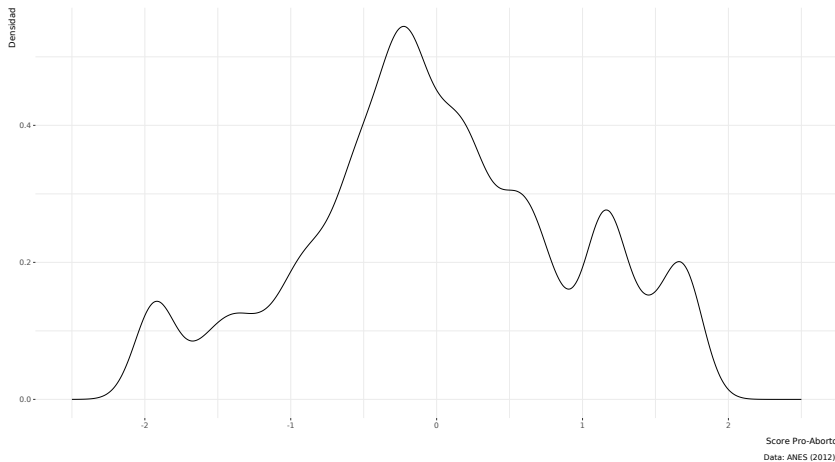
# Nuestros Datos

**Variable Dependiente:** Score pro-aborto.

El estimado tiene media cero y desviación estándar de uno.

- Valores más altos = más “pro-aborto.”

## Densidad del Score Pro-Aborto (ANES, 2012)





# Efectos Interactivos

La fórmula de esta regresión quedaría de la siguiente forma:

$$\hat{y} = \hat{a} + \hat{b}_1(x_1) + \hat{b}_2(x_2) + \hat{b}_3(x_1 * x_2)$$

where:

- $\hat{y}$  = valor estimado de score pro-aborto.
- $x_1$  = partidismo (0 = Dems, 1 = Ind., 2 = Rep).
- $x_2$  = conocimiento político (0 = no conoce al Orador, 1 = conoce al Orador).
- $x_1 * x_2$  = producto de las dos variables.

# Interpretación de los Coeficientes

*Hay que tener cuidado cuando interpretamos los coeficientes de regresión en este caso.*

- El coeficiente de regresión de partidismo es el efecto de partidismo cuando conocimiento político = 0.
- El coeficiente de conocimiento político es el efecto de conocimiento político cuando partidismo = 0 (i.e. entre Demócratas).

```
M5 <- lm(lchoice ~ pid*knowspeaker, data=anes_prochoice)
M5df <- broom::tidy(M5)

stargazer(M1, M2, M4, style="ajps",
  omit.stat=c("F", "ser"), header=FALSE,
  dep.var.labels.include = FALSE,
  covariate.labels = c("partidismo", "conocimiento pol.", "interacción"),
  title= "Efectos Interactivos")
```

**Table 5:** Efectos Interactivos

partidismo	-0.237*** (0.020)
conocimiento político	0.414*** (0.036)
part*conoc	-0.184*** (0.031)
Constant	0.099*** (0.022)
N	5196
R-squared	0.092
Adj. R-squared	0.091

\*\*\*  $p < .01$ ; \*\*  $p < .05$ ; \*  $p < .1$

# Efectos Interactivos

Interpretación de Tabla 5:

- El estimado del score pro-aborto es 0.099 para Demócratas con poco conocimiento político.
- $\hat{b}_1$ ,  $\hat{b}_2$ , and  $\hat{b}_3$  son estadísticamente insignificantes.
- Cuando  $x_1$  y  $x_2 = 1$ , restamos -0.184 de  $\hat{y}$ .
- Conocimiento político produce scores pro-aborto más altos entre demócratas.

# Efectos Interactivos

Esto pasa para demócratas:

- $\hat{y}$  para demócratas con bajo conocimiento: 0.099.
- $\hat{y}$  para demócratas con alto conocimiento: 0.513.

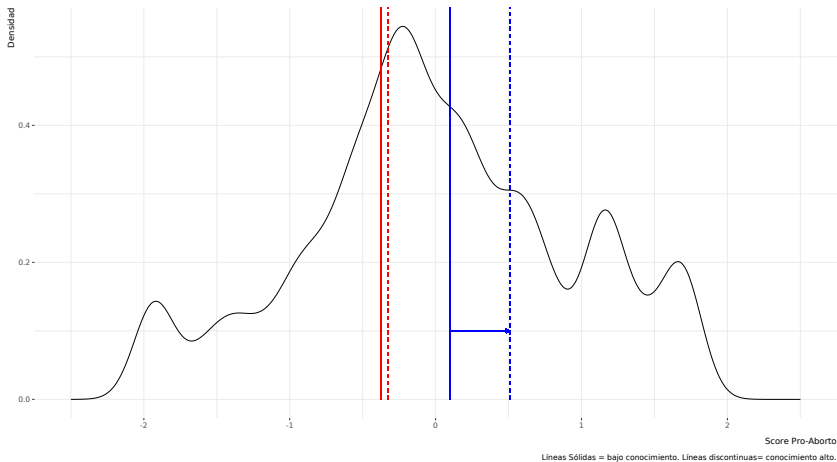
Para Republicanos es más interesante:

- $\hat{y}$  para Republicanos con bajo conocimiento: -0.374.
- $\hat{y}$  para Republicanos con alto conocimiento : -0.328.

Vemos un efecto **grande** de conocimiento político en Demócratas, pero en Republicanos es muy.. **muy pequeño**.

## Plot de Densidad para Score Pro-Aborto con Efectos Interactivos

Hay que notar que el efecto de conocimiento político es mucho más grande para Demócratas que para Republicanos.



# Table of Contents

Introducción

Extendiendo OLS

- Variables Dicotómicas (Dummies)

- Modelos con Efectos Fijos

- Regresión Múltiple

- Efectos Interactivos