Modelos No Lineales (Cameron & Trivedi)

Santiago Alonso-Díaz, PhD

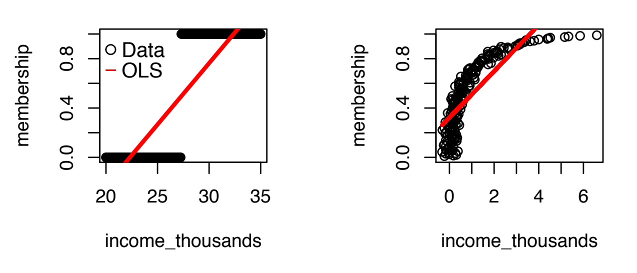
Profesor Asistente
Departamento de Economía
Universidad Javeriana

Modelos No Lineales

Resuelva y escriba la expresión para: Un insecto que pesa 30 gramos al nacer aumenta su peso 20% al mes. ¿Cuál es su peso luego de dos meses?

Modelos No Lineales

Relaciones no lineales no son bien descritas por modelos lineales tradicionales pero si por GLMs.



¿Qué significa el intercepto? La pendiente? ¿OLS captura el salto discreto en la figura de la izq.?

Muchas situaciones no lineales

Table 5.5. *Nonlinear Least Squares: Common Examples*

Model	**Regression Function** $g(\mathbf{x}, \beta)$
Exponential	$\exp(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3)$
Regressor raised to power	$\beta_1 x_1 + \beta_2 x_2^{\beta_3}$
Cobb–Douglas production	$\beta_1 x_1^{\beta_2} x_2^{\beta_3}$
CES production	$[\beta_1 x_1^{\beta_3} + \beta_2 x_2^{\beta_3}]^{1/\beta_3}$
Nonlinear restrictions	$\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$, where $\beta_3 = -\beta_2 \beta_1$

Modelos No Lineales

El estimativo de los parámetros requiere cuidados adicionales. Por ejemplo, si buscamos minimizar cuadrados,

$$\sum_i (y_i - g(x_i, \beta))^2$$

g() es una función lineal (e.g. $x\beta$**) o no lineal (e.g.** $e^(x\beta)$**).**

La derivada (chain rule) la igualamos a cero,

$$2 \times \sum_i \frac{\delta g_i}{\delta \beta}(y_i - g(x_i, \beta)) = 0$$

Si g es lineal i.e. $x\beta$, entonces $\frac{\delta g_i}{\delta \beta} = x_i$
¿Qué significa que el producto (dot) de dos vectores sea cero? Independencia! La función a optimizar (residuales al cuadrado) NO depende de x

Modelos No Lineales

$$2 \times \sum_i \frac{\delta g_i}{\delta \beta}(y_i - g(x_i, \beta)) = 0$$

Si g() es no lineal $\frac{\delta g_i}{\delta \beta}$ no es igual a x i.e. el residual depende de la derivada, no es independiente de x.

No hay forma analítica, se requieren métodos numéricos para estimar $\widehat{\beta}$
(más de esto en siguientes diapositivas)

Modelos No Lineales

Ejemplo método numérico: Gradient Descent
(https://www.youtube.com/watch?v=A6FiCDoz8_4)

Modelos No Lineales

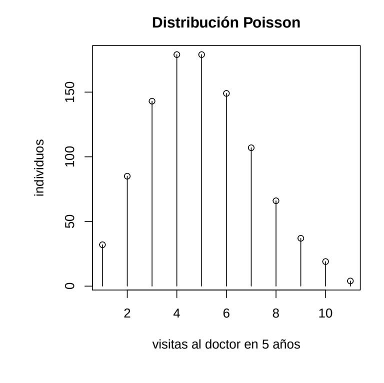
Non_linear_regression.ipynb; Sección Visualización distribuciones y regresión
La visualización esta en Python (o dibujar la regresión (N(BX, sigma))

Ejemplo: Distribución Poisson

Modelos No Lineales

La distribución Poisson sirve para modelar eventos discretos

Por ejemplo, visitas de individuos al doctor en 5 años, número de patentes por mes, o tasa de disparos de una neurona por segundo



La distribución Poisson tiene la siguiente masa,

$$f(y|\lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}$$

y = 0,1,2, ...
λ: parámetro libre
E[y] = λ, Var[y] = λ

Modelos No Lineales

Qué nos interesa en una regresión? $E[y] = x\beta$
Qué es y en un proceso Poisson?

El parámetro λ varía acorde a los predictores y lo definimos,

$$\lambda = e^{x\beta}$$

La masa queda así,

$$f(y|x,\beta) = \frac{e^{-e^{x\beta}}(e^{x\beta})^y}{y!}$$

Modelos No Lineales

Dado que tenemos una función de probabilidad, maximizamos la verosimilitud de los datos para obtener los parámetros β. Si asumimos independencia,

$$\prod_{i}^{n} f(y_i|x_i, \beta)$$

Tomamos el logaritmo por conveniencia computacional (convertir a sumas), usamos el promedio ($\frac{1}{N}$), e igualamos a cero la derivada de $f(y_i|x_i, \beta)$ con respecto a β

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - e^{(x_i\beta)})x_i = 0$$

No hay solución explicita, por lo que el estimativo $\widehat{\beta}$ se obtiene por métodos numéricos

Modelos No Lineales

Interpretación de coeficientes

Modelos No Lineales

Efectos marginales ($\frac{dy}{dx}$). Por ejemplo, cuanto aumenta el número de pacientes con otro seguro médico?. En un modelo lineal es el parámetro β,

$$\frac{dy}{dx}(x\beta) = \beta$$

En un modelo no lineal el β ya no es el efecto marginal. Por ejemplo,

$$\frac{dy}{dx}(e^{x\beta}) = e^{x\beta}\beta$$

El efecto marginal también depende del predictor x

Modelos No Lineales

Como el efecto marginal depende del predictor x, estas son posibles soluciones

Table 5.2. *Marginal Effect: Three Different Estimates*

Formula	**Description**
$N^{-1}\sum_i \partial \mathrm{E}[y_i \vert \mathbf{x}_i]/\partial \mathbf{x}_i$	Average response of all individuals
$\partial \mathrm{E}[y \vert \mathbf{x}]/\partial \mathbf{x}\vert_{\bar{\mathbf{x}}}$	Response of the average individual
$\partial \mathrm{E}[y \vert \mathbf{x}]/\partial \mathbf{x}\vert_{\mathbf{x}^*}$	Response of a representative individual with $\mathbf{x} = \mathbf{x}^*$

Modelos No Lineales

Sección Efectos marginales no lineales dependen del valor del predictor

Modelos No Lineales

Single index models. Si el regresor x_j (e.g. $j = 1$: ingreso, $j = 2 : genes$, etc) y los parámetros β entran a la función no lineal g() directamente i.e. g($x\beta$) por chain rule los efectos marginales son

$$\frac{dy}{dx_j} g(x\beta) = g'(x\beta)\beta_j$$

Modelos No Lineales

Si hacemos una comparación relativa del efecto marginal de dos regresores j & k (e.g. ingreso y genes),

$$\frac{g'(x\beta)\beta_j}{g'(x\beta)\beta_k} = \frac{\beta_j}{\beta_k}$$

Así, por ejemplo, si β_j es el doble de β_k, una unidad de cambio del regresor j (e.g ingreso) tiene el doble de efecto que una unidad del regresor k (e.g. genes)

Modelos No Lineales

En suma, hay que tener cuidado con la interpretación de los coeficientes en un modelo no lineal.

Pueden indicar fuerza de relación con la respuesta y, pero dependen del valor que tome el regresor x.

Estimación de coeficientes (MLE & NLS)

En MLE (maximum likelihood estimate) se maximiza la verosimilitud o likelihood de la data $f(y, X|\theta)$ dado un vector de parámetros θ

OJO: el likelihood function $L_N(\theta|y, X)$ que se usa en MLE es una función de los parámetros.

Modelos No Lineales

NO confundir $f(y, X|\theta)$ (likelihood) con la función $L_N(\theta|y, X)$. Esta última también le dicen likelihood pero es la función a optimizar.

TAMPOCO confundir $L_N(\theta|y, X)$ con posterior bayesiana $f(\theta|y, X)$ que está ponderada por priors. Las funciones f son probabilidades condicionales.

Modelos No Lineales

Sección Diferentes sentidos de likelihood

Para usar MLE necesitamos una densidad i.e. tenemos que parametrizar las densidades en función de los regresores x & parámetros β

Table 5.3. *Maximum Likelihood: Commonly Used Densities*

Model	**Range of** y	**Density** $f(y)$	**Common Parameterization**
Normal	$(-\infty, \infty)$	$[2\pi\sigma^2]^{-1/2}e^{-(y-\mu)^2/2\sigma^2}$	$\mu = \mathbf{x}'\beta, \sigma^2 = \sigma^2$
Bernoulli	0 or 1	$p^y(1-p)^{1-y}$	Logit $p = e^{\mathbf{x}'\beta}/(1+e^{\mathbf{x}'\beta})$
Exponential	$(0, \infty)$	$\lambda e^{-\lambda y}$	$\lambda = e^{\mathbf{x}'\beta}$ or $1/\lambda = e^{\mathbf{x}'\beta}$
Poisson	$0, 1, 2, \ldots$	$e^{-\lambda}\lambda^y/y!$	$\lambda = e^{\mathbf{x}'\beta}$

Modelos No Lineales

Sección MLE

Generalized Linear Models (GLM)
Fuente: [Gelman and Hill, 2006]

Modelos No Lineales

¿Qué es un GLM?

- Predictores $x\beta$
- Una link function g() para la data y e.g. binomial
- g() produce un vector de data transformada $\widehat{y} = g^{-1}(x\beta)$ que se usa para parametrizar la distribución de la data $p(y|\widehat{y})$
- Una distribución de la data $p(y|\widehat{y})$
- Parámetros adicionales (e.g. para la link function o la distribución de la data)

Ejemplo: regresión lineal

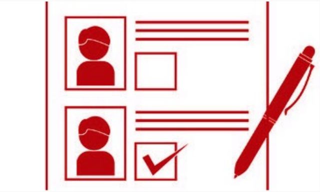
- Predictores: $x\beta$
- Link function: identidad g(u) = u
- Data transformada: $\widehat{y} = g^{-1}(x\beta) = x\beta$
- Distribución data: $y \sim N(\widehat{y}, \sigma)$; σ se estima con la data

Modelos No Lineales

Ejemplo: regresión logística

- Predictores: $x\beta$
- Link function: $g(u) = log\left(\frac{u}{1-u}\right)$
- Data transformada: $\widehat{y} = g^{-1}(x\beta) = \frac{1}{1+e^{-x\beta}}$
- Distribución data: $y \sim binomial(n, p = \widehat{y})$; n es el numero de observaciones

Ejemplo práctico: regresión logistica



La regresión logística se usa para modelar resultados binarios.

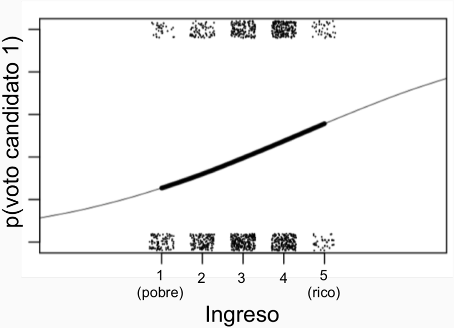
Por ejemplo, el voto por dos candidatos en función del ingreso

Modelos No Lineales

Esta data es de una elección entre dos candidatos. La regresión logística dio $\beta_{intercepto}$ = -1.4, $\beta_{ingreso}$ = 0.33.

Es decir, p(voto a candidato 1) = $logit^{-1}(-1{,}4 + 0{,}33 ingreso) = \frac{1}{1+e^{-(-1{,}4+0{,}33 ingreso)}}$

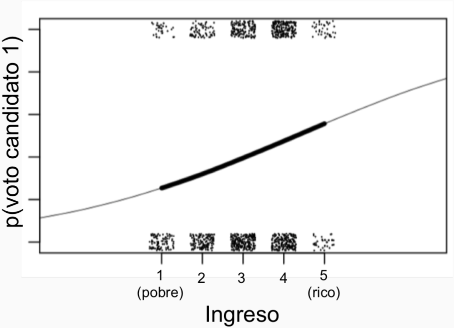
Qué significan los βs? **Con cuidado! no es lineal y la escala ya no es la de la data, todo esta en [0,1].**



Modelos No Lineales

Tip 1: evaluar el modelo en el promedio o media de la data.

Para interpretar $\beta_{intercepto}$ = -1.4:
Ingreso es categórico y la categoría media es 3. Asi,
p(voto a candidato 1) = $\frac{1}{1+e^{-(-1,4+0,33\times 3)}} = 0,4$. Es decir, en el ingreso medio la probabilidad de votar por candidato 1 es 0.4



Modelos No Lineales

Tip 1: evaluar el modelo en el promedio o media de la data.

Para interpretar $\beta_{ingreso} = 0.33$:
Evaluar un cambio de ingreso e.g. ingreso medio - ingreso anterior
$\frac{1}{1+e^{-(-1,4+0,33\times 3)}} - \frac{1}{1+e^{-(-1,4+0,33\times 2)}} = 0{,}08$. Es decir, un aumento de categoría de ingreso aumenta la probabilidad de votar por candidato 1 en 8%

Tip 2: dividir los coeficientes por 4 (excepto el intercepto).

Suena raro pero tiene una razón: $\frac{\beta}{4}$ es la máxima diferencia de cambio de probabilidad en la función logistic correspondiente a un cambio en el predictor x **[Gelman and Hill, 2006]**.

Por ejemplo, $\frac{\beta_{ingreso}}{4} = \frac{0{,}33}{4} = 0{,}08$. Este es el valor de la anterior diapositiva i.e. encontramos el máximo cambio de probabilidad de votar por unidad de ingreso.

OJO: esto es para evaluar cambios máximos y se debe interpretar como tal.

Modelos No Lineales

Tip 3: odds ratios i.e. exponenciar los coeficientes.

$e^{\beta_{ingreso}} = e^{0,33} = 1,39$. Es decir, un cambio en el ingreso aumenta el ratio de posibilidades de votar por candidato 1 en 1.39 (todo lo demás constante).
Qué quiere decir esto? Gelman y colegas afirman, con razón, que los odds ratios son poco intuitivos y prefieren usar cambios en probabilidad como las anteriores diapositivas.

Modelos No Lineales

Qué hacer si hay más regresores? Por ejemplo

p(voto a candidato 1) $= \frac{1}{1+e^{-(-1{,}4+0{,}33 ingreso+2{,}12 educacion)}}$

De nuevo, se puede evaluar el modelo en una referencia e.g. promedio de los otros regresores y evaluar el marginal del regresor de interés

Modelos No Lineales

Ejemplo práctico: Poisson

La distribución Poisson se usa para modelar conteos.

Por ejemplo, número de accidentes en una intersección en función de velocidad y presencia o no de semáforo

Modelos No Lineales

Si y_i es el número de accidentes en la intersección i,

$$y_i \sim Poisson(\theta_i)$$

El parámetro θ_i debe ser positivo, así que

$$\theta_i = e^{x\beta}$$

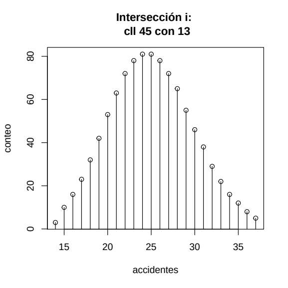
Modelos No Lineales

Para agregar un baseline o exposure en los conteos (e.g. para controlar por número de vehículos que pasan por la intersección) añadimos un parámetro,

$$y_i \sim Poisson(exposure_i \times \theta_i)$$

El log(exposure) se llama offset

Modelos No Lineales



Suponga que luego de ajustar el modelo encuentra que,

$$y_i \sim Poisson(e^{2{,}8+0{,}012AvgVel_i-0{,}2TfcLight_i})$$

Qué significan los βs de los regresores?

Modelos No Lineales

$$y_i \sim Poisson(e^{2,8+0,012AvgVel_i-0,2TfcLight_i})$$

Para AvgVel, el coeficiente 0.012 es la diferencia esperada en accidentes por cada km/h adicional.

por qué? $E[y] = \theta$ para una Poisson, por eso podemos interpretar $e^{0,012} = 1{,}012$ como un incremento de 1.2% en accidentes por cada km/h adicional.

-0.2 para traffic light (0: no hay semáforo, 1: si)? Es similar, use el signo y la magnitud $e^{-0,2}$

Modelos No Lineales

Paréntesis conceptual: cuando Binomial o Poisson?

- Binomial se usa cuando la data y_i son eventos sobre un total n
- Poisson se usa cuando la data y_i son eventos sin un límite n superior

Modelos No Lineales

Sección Poisson police stops

Modelos No Lineales

En este punto debe ser claro que regresión no-lineal difiere de la lineal (estimación e interpretación). Reforcemos con el siguiente ejemplo,

$$y = \lambda e^{-\lambda}$$

$\lambda = e^{(\beta_1 + \beta_2 x)}$, y asumamos que $\beta_1 = 2$ y $\beta_2 = -1$

Modelos No Lineales

Todos los métodos no lineales aciertan los βs pero difieren en OLS Por qué? Identificación

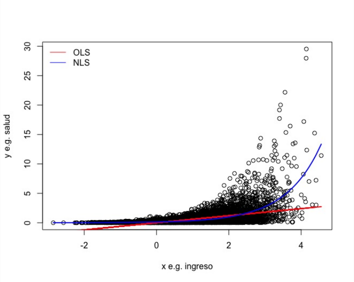
Imagine que y es salud, x ingreso. Signos en OLS diferentes conclusiones diferentes? NO!!

Table 5.7. *Exponential Example: Least-Squares and ML Estimates*[a]

	Estimator				
Variable	**OLS**	**ML**	**NLS**	**WNLS**	**FGNLS**
Constant	−0.0093	1.9829	1.8876	1.9906	1.9840
	(0.0161)	(0.0141)	(0.0307)	(0.0225)	(0.0148)
	[0.0172]	[0.0144]	[0.1421]	[0.0359]	[0.0146]
			{0.2110}		
x	0.6198	−0.9896	−0.9575	−0.9961	−0.9907
	(0.0113)	(0.0099)	(0.0097)	(0.0098)	(0.0100)
	[0.0254]	[0.0099]	[0.0612]	[0.0224]	[0.0101]
			{0.0880}		
lnL	–	−208.71	−232.98	−208.93	−208.72
R^2	0.2326	0.3906	0.3913	0.3902	0.3906

Modelos No Lineales

Los (signos) β en regresiones no lineales NO se pueden interpretar como efectos marginales i.e. y es monotónica en x tanto para OLS y NLS



Calcule la derivada dy/dx de $y = e^{(-(\beta_1+\beta_2 x))}$ (chain rule). Es positivo en todo el dominio?

Caso probit: cyberbullying CAUSA aumento de tasas de suicidio?

[Nikolaou, 2017]

Algunos datos de Cyberbullying en Colombia. Lo sufren más los hombres y en colegios públicos

Tabla 3 Implicación en *bullying* (%)

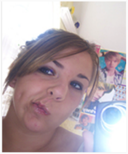
	Sexo		Edad			Estrato				Colegio		Curso					
Implicación	Masc. n = 878	Femen. n = 999	≤ 13 n = 709	14-15 n = 663	≥ 16 n = 506	Bajo n = 400	Medio-bajo n = 509	Medio-alto n = 608	Alto n = 339	Público n = 1253	Privado n = 629	6.º n = 323	7.º n = 348	8.º n = 379	9.º n = 301	10.º n = 237	11.º n = 293
Victimización	53.6***	46.4	42.9	33.1	24	20.8	25.5	35.2**	18.5	62.6***	37.4	19.3	21.5	18.4	14.7	10.9	15.2
Agresión	67.1***	32.9	34.1	34.1	31.8	21.4	25	32.2**	21.4	62.4***	37.6	7.1	20	23.5***	15.3	15.3	18.8
Agresión-victimización	61.2***	38.8	32.4	42.1*	25.5	17.7	22.4	33.5**	26.4	53.5***	46.5	10.4	15.8	24.6**	19.2	15	15

* $p < .05$.
** $p < .01$.
*** $p < .001$.

[Herrera-Lopez et al., 2017]

La conexión con suicidio es anecdótica o correlacional. Algunos casos mediáticos en EE.UU.

Ryan Halligan
1989-2003
Niña fingio que lo quería
para reirse de sus chats



Megan Meier
1992-2006
Crearon una cuenta
falsa para enamorarla

Modelos No Lineales

Antecedentes/literatura previa. Problemática (económica no moral).

- Cyberbullying puede reducir el capital de salud individual (intentos de suicidio fallidos)
- También el capital de salud de la economía (intentos de suicidio fatales)

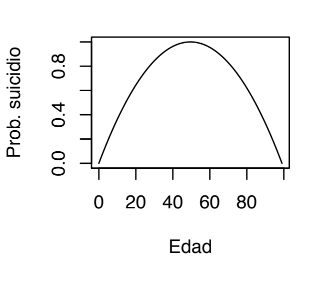
[Nikolaou, 2017]

Antecedentes/literatura previa. Relación con economía.

- Tiene una racionalidad si el valor futuro de mantenerse vivo es negativo
- Mayor probabilidad de suicidio si el flujo de ingresos futuros es bajo o incierto
- Hay más suicidios en recesiones económicas pero en términos relativos i.e. si todos están desempleados no hay más suicidios.

[Nikolaou, 2017]

Antecedentes/literatura previa. Relación con edad (gráfica ilustrativa, no datos reales)



[Nikolaou, 2017]

Modelos No Lineales

Antecedentes/literatura previa. Relación con género

- Mujeres tienen más ideaciones de suicidio no fatal
- Los hombres completan más intentos de suicidio.
- Mujeres escogen métodos menos violentos (pastillas) que los hombres (pistolas u colgarse).

[Nikolaou, 2017]

Modelos No Lineales

Antecedentes/literatura previa. Relación con variables socioeconómicas

- Raza: blancos caucásicos se suicidan más
- Salud física (e.g. obesidad) y mental (depresión, anorexia) lleva a más suicidios.
- La correlación entre nivel de ingresos y tasa de suicidios es baja.

[Nikolaou, 2017]

Modelos No Lineales

Antecedentes/literatura previa. Relación con juventud y cyberbullying

- Hay muchos estudios correlacionales que conectan suicidio y cyberbullying.
- Estudios causales previos muestran que bullying perjudica resultados educativos y laborales.
- Ninguno conecta causalmente cyberbullying con suicidio

[Nikolaou, 2017]

Modelo cognitivo. El objetivo de todo joven es maximizar Z.

$$Z(a,Y,s) = \int_a^{\omega} e^{-r(m-a)} U[C(m, Y(s(\cdot))) - K(m, s(\cdot))] P(m, s(\cdot)) dm$$

U(), C(), & K() son utilidad, consumo, y costos de mantenerse vivo para cada edad m, respectivamente.

P() probabilidad de sobrevivir hasta la edad m condicional a que se vive hasta a.

Y nivel de ingreso disponible para consumir.

ω **expectativa de vida máxima.**

r tasa de descuento

s() función de comportamiento suicida.

[Nikolaou, 2017]

Modelos No Lineales

Tres aspectos relevantes de la formula: 1) ideación suicida (s()) es input de la función P(), probabilidad de sobrevivir. Es decir, en el modelo tener ideas suicidas reduce la probabilidad de seguir vivo.

$$Z(a, Y, s) = \int_a^{\omega} e^{-r(m-a)} U[C(m, Y(s(\cdot))) - K(m, s(\cdot))] P(m, s(\cdot)) dm$$

[Nikolaou, 2017]

Modelos No Lineales

Tres aspectos relevantes de la formula: 2) C, consumo, y K, mantenerse con vida, también dependen de ideaciones suicidas (s()). Por ejemplo, si una ideación suicida se ejecuta fallidamente, eso puede provocar verguenza y disminuir consumo (e.g. no salir con amigos después del intento).

$$Z(a, Y, s) = \int_a^{\omega} e^{-r(m-a)} U[C(m, Y(s(\cdot))) - K(m, s(\cdot))] P(m, s(\cdot)) dm$$

[Nikolaou, 2017]

Modelos No Lineales

Tres aspectos relevantes de la formula: 3) la función s(), ideación suicida, depende de características individuales X, variables no observadas u, y la calidad del ambiente social Q del estudiante. La derivada parcial con respecto a Q es negativa i.e mejor ambiente Q menor ideación suicida.

$$s = s(X, Q; u), \frac{ds}{dQ} < 0$$

[Nikolaou, 2017]

Modelos No Lineales

Por otro lado, $\frac{dQ}{dCyberbullying} < 0$ i.e. cyberbullying reduce calidad de ambiente social Q.

También se asume en el modelo que $\frac{dCyberbullying}{dLaw} < 0$ i.e. leyes pueden reducir cyberbullying.

[Nikolaou, 2017]

La decisión final de suicidarse o no es una comparación entre el valor esperado Z dado que hay suicidio (s()=1) o no (s()=0).

$$Z(\boldsymbol{a}, Y|s(\,\cdot\,) = 1) > Z(\boldsymbol{a}, Y|s(\,\cdot\,) = 0)$$

[Nikolaou, 2017]

Modelos No Lineales

Ahora sí las regresiones. Data=Youth Risk Behavioral Survey (1991-2015)
El instrumento propuesto es leyes en contra de cyberbullying.
Discusión Es un buen instrumento bajo el modelo propuesto? Las leyes afectan directamente el comportamiento suicida?

Formula de regresión probit. Segunda etapa:

$$S_{it} = \beta_0 + \beta_1 C_{it} + \beta_2 X_{it} + \beta_3 Z_{st} + \beta_4 H_{st} + \mu_s + \gamma_t + u_{it}$$

Primera etapa (instrumentos):

$$C_{it} = \alpha_0 + \alpha_1 CBL_{st} + \alpha_2 X_{it} + \alpha_3 Z_{st} + \alpha_4 H_{st} + \mu_s + \gamma_t + \epsilon_{it}$$

S_{it}: comportamiento suicida individuo i en periodo t; C_{it}: cyberbullying; X_{it}: características socioeconómicas; Z_{st}: características económicas del estado s en periodo t; $H_S t$ controles de métricas de salud; μ_S: efectos fijos de estado; γ_t: shocks en tasa de suicidio; μ_{it}: shocks en preferencias de suicidio; CBL_{st}: leyes combatiendo cyberbullying en estado s en periodo t. [Nikolaou, 2017]

Modelos No Lineales

Por qué nos interesa más β_1 que α_1?

$$S_{it} = \beta_0 + \beta_1 C_{it} + \beta_2 X_{it} + \beta_3 Z_{st} + \beta_4 H_{st} + \mu_s + \gamma_t + u_{it}$$

$$C_{it} = \alpha_0 + \alpha_1 CBL_{st} + \alpha_2 X_{it} + \alpha_3 Z_{st} + \alpha_4 H_{st} + \mu_s + \gamma_t + \epsilon_{it}$$

[Nikolaou, 2017]

Modelos No Lineales

Por qué instrumentalizar la variable cyberbullying C_{it}? Es endógena? Pertenecer a la población cyberbullied es aleatorío? O los cyberbullied tienen características especiales?

$$S_{it} = \beta_0 + \beta_1 C_{it} + \beta_2 X_{it} + \beta_3 Z_{st} + \beta_4 H_{st} + \mu_s + \gamma_t + u_{it}$$

$$C_{it} = \alpha_0 + \alpha_1 CBL_{st} + \alpha_2 X_{it} + \alpha_3 Z_{st} + \alpha_4 H_{st} + \mu_s + \gamma_t + \epsilon_{it}$$

[Nikolaou, 2017]

Efectos marginales (2nd stage): Cyberbullying aumenta 11% planes, 8.7% intentos, y 4.8% heridas por suicidio.

Marginal effects of c

	Second stage		
	Suicide plans [3]	Attempt suicide [4]	Suicide injury [5]
Cyberbullying	0.117***	0.087***	0.048***
	(0.008)	(0.030)	(0.009)
Male	−0.057***	−0.008**	0.001
	(0.004)	(0.003)	(0.001)
Black	−0.021***	0.018***	0.012***
	(0.003)	(0.005)	(0.002)
Hispanic	0.003	0.021***	0.010***
	(0.004)	(0.003)	(0.002)
Other race	0.018***	0.020***	0.009***
	(0.002)	(0.002)	(0.001)
Age 15	0.005***	0.005***	0.004***
	(0.001)	(0.002)	(0.001)
Age 16	0.013***	0.014***	0.008***
	(0.002)	(0.002)	(0.001)
Age 17	0.016***	0.018***	0.011***
	(0.002)	(0.003)	(0.002)
Age 18 plus	0.013***	0.024***	0.015***
	(0.003)	(0.004)	(0.002)
Grade 10	−0.008***	−0.008***	−0.004***
	(0.002)	(0.001)	(0.001)
Grade 11	−0.021***	−0.019***	−0.008***
	(0.002)	(0.002)	(0.001)
Grade 12	−0.029***	−0.027***	−0.011***
	(0.003)	(0.003)	(0.002)
Underweight	0.011***	0.003	0.000
	(0.004)	(0.002)	(0.001)
Overweight	0.012***	0.007***	0.003***

Modelos No Lineales

Efectos marginales (1st stage): La leyes antibullying y anticyberbullying disminuyen cyberbullying pero no pensamientos suicidas buen instrumento?

Marginal effects of cyberbullying on suicidal behaviors, bivariate probit estimates.

	First stage	
	Cyberbullying [1]	Suicidal thoughts [2]
Obese	0.010***	0.033***
	(0.003)	(0.003)
Unsafe at school	0.174***	0.146***
	(0.006)	(0.005)
Unemployment rate	0.001	−0.008
	(0.002)	(0.006)
Beer tax	0.002	−0.005
	(0.006)	(0.012)
Cigarettes tax	−0.000	0.000
	(0.000)	(0.000)
Population	−0.000	0.000
	(0.001)	(0.001)
Mental health expenses	−0.000	0.000
	(0.000)	(0.000)
Health consumption	0.000	0.000
	(0.000)	(0.000)
Suicide prevention law	−0.000	0.019**
	(0.004)	(0.009)
Mental health law	0.007*	0.012*
	(0.004)	(0.007)
Anti-bullying law	−0.015**	−0.005
	(0.006)	(0.009)
Income (per capita)	−0.000	−0.000
	(0.000)	(0.000)
Cyberbullying law	−0.010**	–

Modelos No Lineales

Estamos seguros que encontramos un efecto causal via el instrumento?
Comparamos niños con las mismas características?

Paréntesis pedagógico: Propensity Score Matching

Modelos No Lineales

Objetivo: construir grupos de control con la data que se tiene (que no siempre viene de experimentos propiamente diseñados)

Modelos No Lineales

Opción 1: Exact matching.
Ideal pero problema combinatorio que explota rápidamente con número de características a match.

Figure 8.1 Exact Matching on Four Characteristics

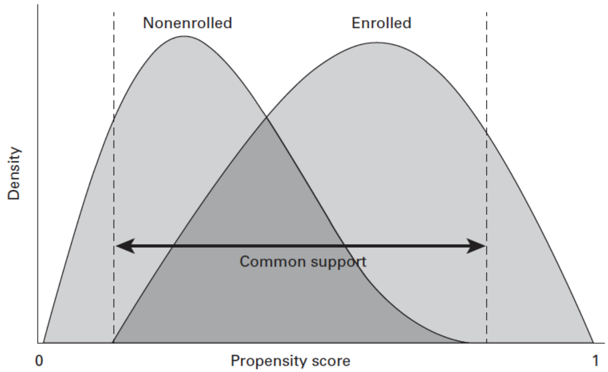
Treated units			
Age	Gender	Months unemployed	Secondary diploma
19	1	3	0
35	1	12	1
41	0	17	1
23	1	6	0
55	0	21	1
27	0	4	1
24	1	8	1
46	0	3	0
33	0	12	1
40	1	2	0

Untreated units			
Age	Gender	Months unemployed	Secondary diploma
24	1	8	1
38	0	1	0
58	1	7	1
21	0	2	1
34	1	20	0
41	0	17	1
46	0	9	0
41	0	11	1
19	1	3	0
27	0	4	0

[Gertler et al., 2016]

Opción 2: Propensity score matching.
A cada unidad/individuo le asignamos una probabilidad de ser tratado basado en las características. Es una medida continua entre 0-1.

Figure 8.2 Propensity Score Matching and Common Support



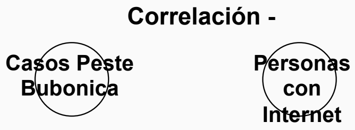
[Gertler et al., 2016]

Modelos No Lineales

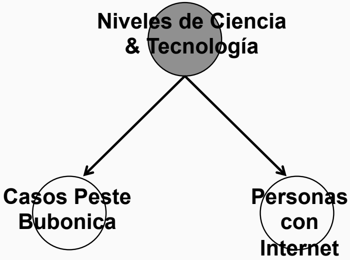
Supuestos para hacer un buen matching

- Independencia condicional

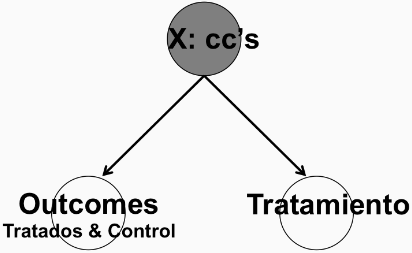
Recordemos independencia condicional



Recordemos independencia condicional



En matching, necesitamos que el tratamiento (e.g. ayuda monetaria o no) sea independiente de los resultados (e.g. vocabulario) dadas una características (e.g. nivel de ingreso familiar).
Por ejemplo, un programa de becas que ayude monetariamente no vaya a priori a buenos estudiantes si el objetivo es ver mejora en el vocabulario.



Modelos No Lineales

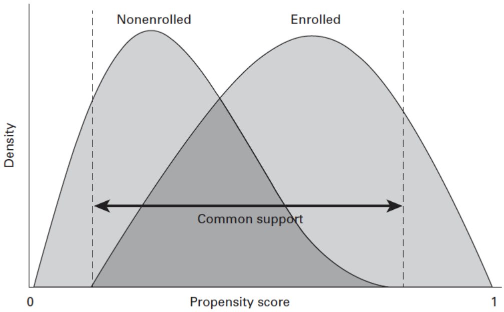
Supuestos para hacer un buen matching

- Independencia condicional
- Soporte común

Modelos No Lineales

Soporte común:
$0 < P[tratamiento = 1|x] < 1$ i.e. la probabilidad de ser tratado, dada unas características x, es mayor a cero.
Por qué necesitamos esto? Qué pasa si las dos curvas de la gráfica no se intersectan?

Figure 8.2 Propensity Score Matching and Common Support



[Gertler et al., 2016]

Modelos No Lineales

Cómo calcular propensity matching score? i.e.
p(tratamiento|características)

Una forma es con una regresión logística i.e. $tratamiento \sim caracteristicas$.

Todos los que tengan valor logit suficientemente cercano agruparlos.

Modelos No Lineales

Sección Propensity score matching

Modelos No Lineales

Volvamos a cyberbullying ... en las siguientes diapositivas con la data PSM (propensity score match)

Marginal effects of c

	Second stage		
	Suicide plans [3]	Attempt suicide [4]	Suicide injury [5]
Cyberbullying	0.117***	0.087***	0.048***
	(0.008)	(0.030)	(0.009)
Male	−0.057***	−0.008**	0.001
	(0.004)	(0.003)	(0.001)
Black	−0.021***	0.018***	0.012***
	(0.003)	(0.005)	(0.002)
Hispanic	0.003	0.021***	0.010***
	(0.004)	(0.003)	(0.002)
Other race	0.018***	0.020***	0.009***
	(0.002)	(0.002)	(0.001)
Age 15	0.005***	0.005***	0.004***
	(0.001)	(0.002)	(0.001)
Age 16	0.013***	0.014***	0.008***
	(0.002)	(0.002)	(0.001)
Age 17	0.016***	0.018***	0.011***
	(0.002)	(0.003)	(0.002)
Age 18 plus	0.013***	0.024***	0.015***
	(0.003)	(0.004)	(0.002)
Grade 10	−0.008***	−0.008***	−0.004***
	(0.002)	(0.001)	(0.001)
Grade 11	−0.021***	−0.019***	−0.008***
	(0.002)	(0.002)	(0.001)
Grade 12	−0.029***	−0.027***	−0.011***
	(0.003)	(0.003)	(0.002)
Underweight	0.011***	0.003	0.000
	(0.004)	(0.002)	(0.001)
Overweight	0.012***	0.007***	0.003***

[Nikolaou, 2017]

Distribución PSM

Por qué se cumple el supuesto de common support?

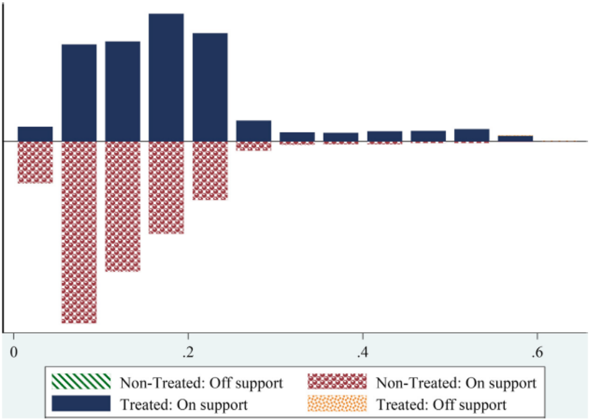


Fig. 1. Propensity score distributions for treated and non-treated students.

[Nikolaou, 2017]

Limites de propensity score matching

- Se requiere una base de datos grande que cumpla el supuesto de common support
- El match se hace solo con observables i.e. asume que los no observables en las unidades control y tratamiento son similares

El resultado sobre cyberbullying y tendencias suicidas se mantienen con distintos algoritmos PSM

Table 3
Propensity score estimates for the effects of cyberbullying on suicidal behaviors.

	Suicidal thoughts	Suicide plans	Attempt suicide	Suicide injury
Nearest neighbor	0.149***	0.118***	0.071***	0.023***
	(0.004)	(0.004)	(0.004)	(0.003)
Radius	0.179***	0.139***	0.091***	0.032***
	(0.007)	(0.007)	(0.005)	(0.003)
Kernel	0.143***	0.110***	0.064***	0.024***
	(0.006)	(0.004)	(0.003)	(0.002)
Mean standardized bias (MSB)[a]	0.7	0.7	0.7	0.7
Pseudo-R^2[a,b]	0.001	0.001	0.001	0.001

[Nikolaou, 2017]

Modelos No Lineales

Es cyberbullying law un buen instrumento para cyberbullying? Endogeneidad: Se relaciona la ley con cyberbullying? (panel A incluye # de estados cercanos con leyes; para controlar spillover effects) Falsificación; Se relaciona la ley con otras variables/unobserved? (panel B)

Table 4
Additional tests for the validity of cyberbullying laws as instruments for cyberbullying in the bivariate probit models.

Panel A—policy endogeneity[a]

	Cyberbullying law	Cyberbullying
Number of neighboring states	0.519***	–
	(0.115)	–
Cyberbullying law	–	–0.010*
	–	(0.004)

Panel B—falsification tests[b]

	Violent crime	Property crime	Property crime
Cyberbullying law	–0.031	–0.023	–0.002
	(0.182)	(0.090)	(0.008)

[Nikolaou, 2017]

Modelos No Lineales

Modelos multinomiales

Sección Multinomial M and M

Modelos No Lineales

Otros ejemplos:

- Transporte (bus, carro, bicicleta, taxi, caminar)
- Seguro médico (EPS, EPS-Prepagada, Ninguno)
- Estatus laboral (Tiempo completo, Medio tiempo, Desempleado)
- Ocupación (Ingeniero, Músico, Médico, Economista, Psicólogo, Abogado)

Multinomial logit (con categorías que NO se pueden ordenar)

Modelos No Lineales

¿Cuál color de m&m escogen?

¿Cuál consola escogen?

Additive random-utility model:
Con categorías multinomiales que no tienen un orden claro, se asume que los individuos escogen la que les represente mayor utilidad

La utilidad de la alternativa j para el individuo i es

$$U_{ij} = Valor_{ij} + \epsilon_{ij}$$

Valor depende de x: aspectos específicos de la alternativa (e.g. precio), & z: aspectos independientes de la alternativa (e.g. genero del individuo)

$$Valor_{ij} = x_{ij}\beta + z_i\gamma_j$$

La alternativa escogida (y) es la de mayor utilidad

$$P(y = j) = P(U_{ij} \geq U_{ik}) para\, todo\; k \neq j$$

Modelos No Lineales

Una forma de asignar utilidades en función de regresores j es con la función softmax,

$$p_{ij} = \frac{e^{x_i \beta_j}}{\sum_{l=1}^{m} e^{x_i \beta_l}}$$

Los βs incluyen el intercepto: los otros coeficientes son relativos a ese valor como en un regresión lineal clásica

Se asume que el individuo i escoge la opción j con probabilidad dada por la formula i.e $\sum_{j=1}^{m} p_{ij} = 1$

Sección Regresión Multinomial (non-ordinal logistic regression)

Multinomial logit (con categorías que se pueden ordenar)

Modelos No Lineales

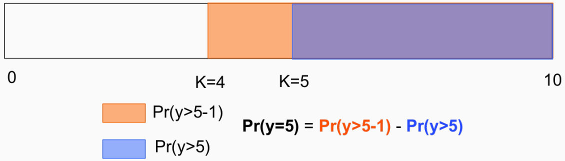
Las K categorías se pueden modelar en función de una variable continua $X\beta$. Los puntos de cortes c son estrictamente crecientes i.e. $0 = c_1 < c_2 < ... < c_k$

$$
\begin{aligned}
P(y > 1) &= logit^{-1}(X\beta - c1) \\
P(y > 2) &= logit^{-1}(X\beta - c2) \\
P(y > 3) &= logit^{-1}(X\beta - c3) \\
&... \\
P(y > K - 1) &= logit^{-1}(X\beta - c_{k-1})
\end{aligned}
$$

La resta de dos valores consecutivos nos da P(y=k),

$$
\begin{aligned}
P(y = k) &= P(y > k - 1) - P(y > k) \\
P(y = k) &= logit^{-1}(X\beta - c_{k-1}) - logit^{-1}(X\beta - c_k)
\end{aligned}
$$

Visualización de $P(y = k)$ i.e. la porción naranja que queda descubierta



Otra forma de escribirlo con una variable latente z:

$$f(x) = \begin{cases} 1 & \textit{if } z_i < 0 \\ 2 & \textit{if } z_i \in (0, c_2) \\ 3 & \textit{if } z_i \in (c_2, c_3) \\ ... & \\ K-1 & \textit{if } z_i \in (c_{k-2}, c_{k-1}) \\ K & \textit{if } z_i > c_{k-1} \end{cases}$$

$$z_i = X_i \beta + \epsilon_i$$

Con distribución logística para el error ϵ_i i.e. $\frac{1}{1+e^{-\epsilon_i}}$. Hay que estimar β y los cortes c (con MLE).

Ejemplo (experimento):

6 jugadores van a votar sobre dos temas durante 30 rondas

Cada jugador tiene tres votos y cada ronda gana el tema con más votos. Primero se vota por tema 1 y los votos restantes van automáticamente al tema 2.

Antes del voto por el tema 1, los experimentadores le dicen a los sujetos que tanto ganan si pasa el tema en esa ronda (entre 1-100).

El número de votos (categorías) y_i se pueden modelar así. El σ se puede pensar como fuzzyness de las categorías & z como utilidad esperada.

$$y_i = \begin{cases} 1 & if\ z_i < c_{1|2} \\ 2 & if\ z_i \in (c_{1|2}, c_{2|3}) \\ 3 & if\ z_i > c_{2|3} \end{cases}$$

$$z_i \sim logistic(ganancia_i, \sigma^2)$$

- Gelman, A. and Hill, J. (2006).
  ***Data analysis using regression and multilevel/hierarchical models*.**
  Cambridge university press.
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., and Vermeersch, C. M. (2016).
  ***Impact evaluation in practice*.**
  The World Bank.
- Herrera-Lopez, M., Romera, E., and Ortega-Ruiz, R. (2017).
  Bullying y cyberbullying en colombia; coocurrencia en adolescentes escolarizado.
  Revista Latinoamericana de Psicologia, 49:163–172.

Nikolaou, D. (2017).
Does cyberbullying impact youth suicidal behaviors?
Journal of health economics, 56:30–46.