

Funciones de Regresión No Lineales

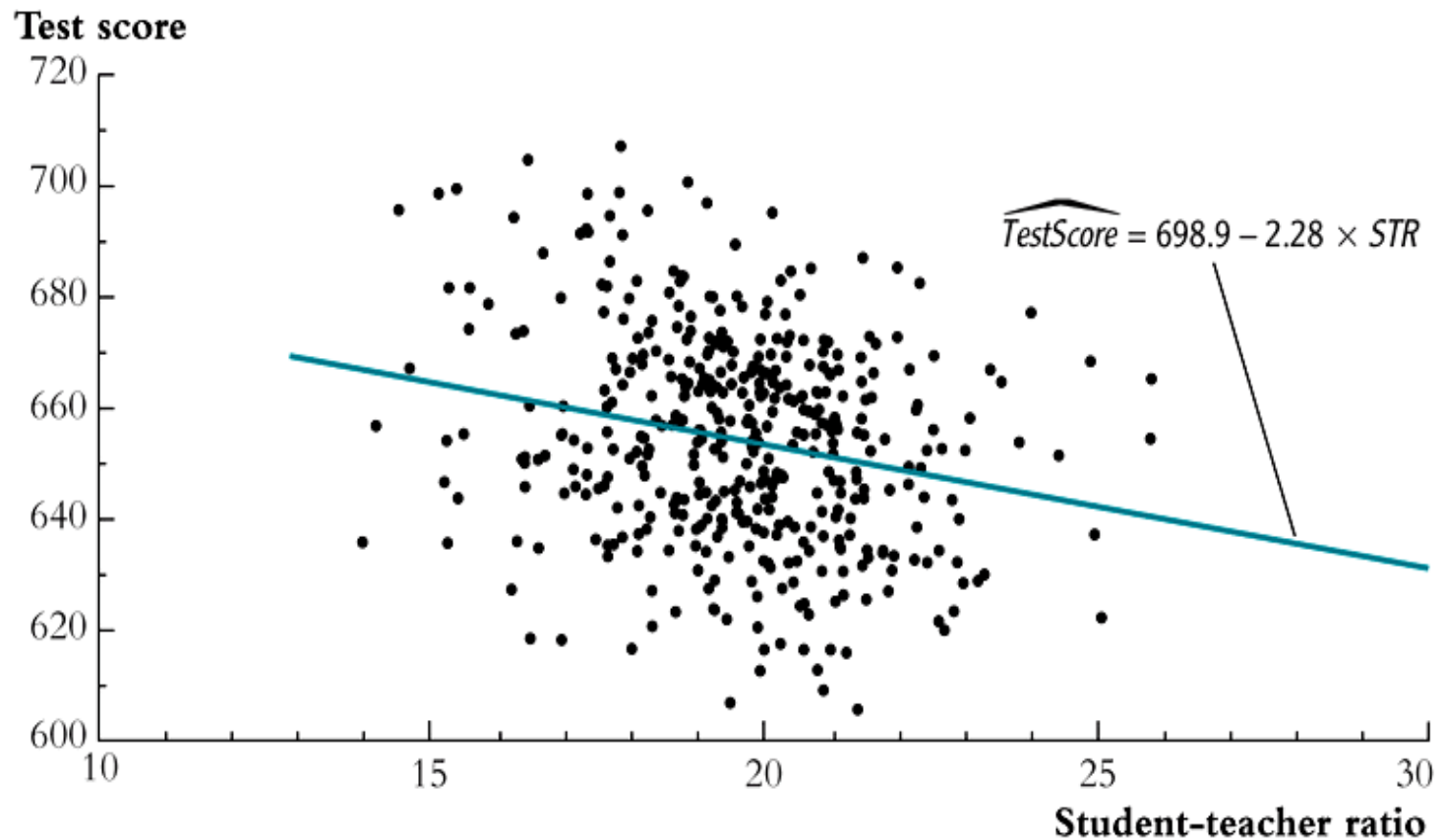
(SW Cap. 6)

- Todo anteriormente ha sido lineal en las X 's
- La aproximación de que la función de regresión es lineal puede ser satisfactoria para algunas variables pero no para otras.
- El esquema de regresión múltiple se puede extender para tener en cuenta funciones de regresión que no son lineales en una o más de una X .

La relación *Notas* – *STR* parece aproximadamente lineal...

FIGURE 4.3 The Estimated Regression Line for the California Data

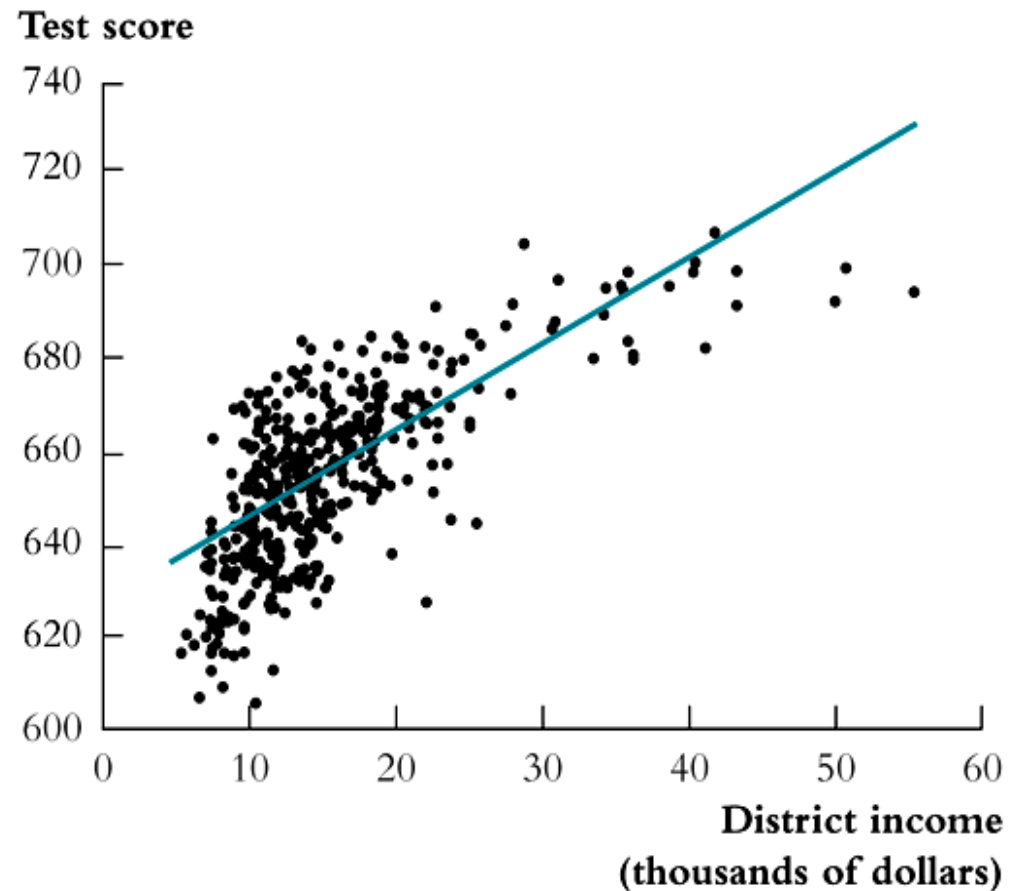
The estimated regression line shows a negative relationship between test scores and the student-teacher ratio. If class sizes fall by 1 student, the estimated regression predicts that test scores will increase by 2.28 points.



Pero la relación entre *Notas* – la renta media del distrito parece como si fuese no lineal.

FIGURE 6.2 Scatterplot of Test Score vs. District Income with a Linear OLS Regression Function

There is a positive correlation between test scores and district income (correlation = 0.71), but the linear OLS regression line does not adequately describe the relationship between these variables.



Si una relación entre Y y X es no lineal:

- El efecto sobre Y de un cambio en X depende del valor de X – es decir, el efecto marginal de X no es constante
- Una regresión lineal está mal especificada – la forma funcional no es la correcta
- El estimador del efecto en Y de X es sesgado – no necesita aún ser correcto en media.
- La solución a este problema es estimar una función de regresión que no sea lineal en X

La función de regresión poblacional no lineal general

$$Y_i = f(X_{1i}, X_{2i}, \dots, X_{ki}) + u_i, i = 1, \dots, n$$

Hipótesis

1. $E(u_i | X_{1i}, X_{2i}, \dots, X_{ki}) = 0$ (igual); implica que f es la esperanza condicionada de Y dadas las X 's.
2. $(X_{1i}, \dots, X_{ki}, Y_i)$ son i.i.d. (igual).
3. existen “bastantes más” momentos (misma idea; la afirmación precisa depende de la f específica).
4. No hay multicolinealidad perfecta (misma idea; la afirmación precisa depende de la f específica).

El efecto esperado sobre Y de un cambio en X_1 en el modelo de Regresión No Lineal (6.3)

El cambio esperado en Y , ΔY , asociado con el cambio en X_1 , ΔX_1 , manteniendo constantes X_2, \dots, X_k , es la diferencia entre el valor de la función de regresión poblacional antes y después de cambiar X_1 , manteniendo X_2, \dots, X_k constantes. Es decir, el cambio esperado en Y es la diferencia:

$$\Delta Y = f(X_1 + \Delta X_1, X_2, \dots, X_k) - f(X_1, X_2, \dots, X_k). \quad (6.5)$$

El estimador de esta diferencia poblacional desconocida es la diferencia entre los valores predichos para estos dos casos. Sea $\hat{f}(X_1, X_2, \dots, X_k)$ el valor predicho de el valor predicho de Y basado en el estimador \hat{f} de la función de regresión poblacional. Entonces, el cambio predicho en Y es

$$\Delta\hat{Y} = \hat{f}(X_1 + \Delta X_1, X_2, \dots, X_k) - \hat{f}(X_1, X_2, \dots, X_k) \quad (6.6)$$

Funciones no lineales de una variable independiente única (SW Sección 6.2)

Veremos dos aproximaciones complementarias:

1. Polinomios en X

La función de regresión poblacional se aproxima por un polinomio cuadrático, cúbico, o de mayor orden.

2. Transformaciones Logarítmicas

- Y y/o X se transforman tomando sus logaritmos
- Esto proporciona una interpretación de “porcentajes” que tiene mucho en muchas aplicaciones

1. Polinomios en X

Se aproxima la función de regresión poblacional por un polinomio:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_r X_i^r + u_i$$

- Este es justo el modelo de regresión lineal múltiple – excepto por el hecho de que los regresores son potencias de X !
- La estimación, el contraste de hipótesis, etc. se realiza como en el modelo de regresión múltiple usando OLS.
- Los coeficientes son difíciles de interpretar, pero la función de regresión por sí misma es interpretable

Ejemplo: la relación Notas – renta

$Renta_i$ = renta media por distrito en el distrito i -ésimo
(miles de dólares per cápita)

Especificación cuadrática:

$$Notas_i = \beta_0 + \beta_1 Renta_i + \beta_2 (Renta_i)^2 + u_i$$

Especificación cúbica:

$$Notas_i = \beta_0 + \beta_1 Renta_i + \beta_2 (Renta_i)^2 \\ + \beta_3 (Renta_i)^3 + u_i$$

Estimación de la especificación cuadrática en STATA

```
generate avginc2 = avginc*avginc;  
reg testscr avginc avginc2, r;
```

Create a new regressor

Regression with robust standard errors

Number of obs = 420
F(2, 417) = 428.52
Prob > F = 0.0000
R-squared = 0.5562
Root MSE = 12.724

testscr	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
avginc	3.850995	.2680941	14.36	0.000	3.32401	4.377979
avginc2	-.0423085	.0047803	-8.85	0.000	-.051705	-.0329119
_cons	607.3017	2.901754	209.29	0.000	601.5978	613.0056

Es estadístico t de $Renta^2$ es -8.85, de manera que la hipótesis de linealidad se rechaza frente a la alternativa cuadrática para un nivel de significación del 1%.

Interpretación de la función de regresión estimada:

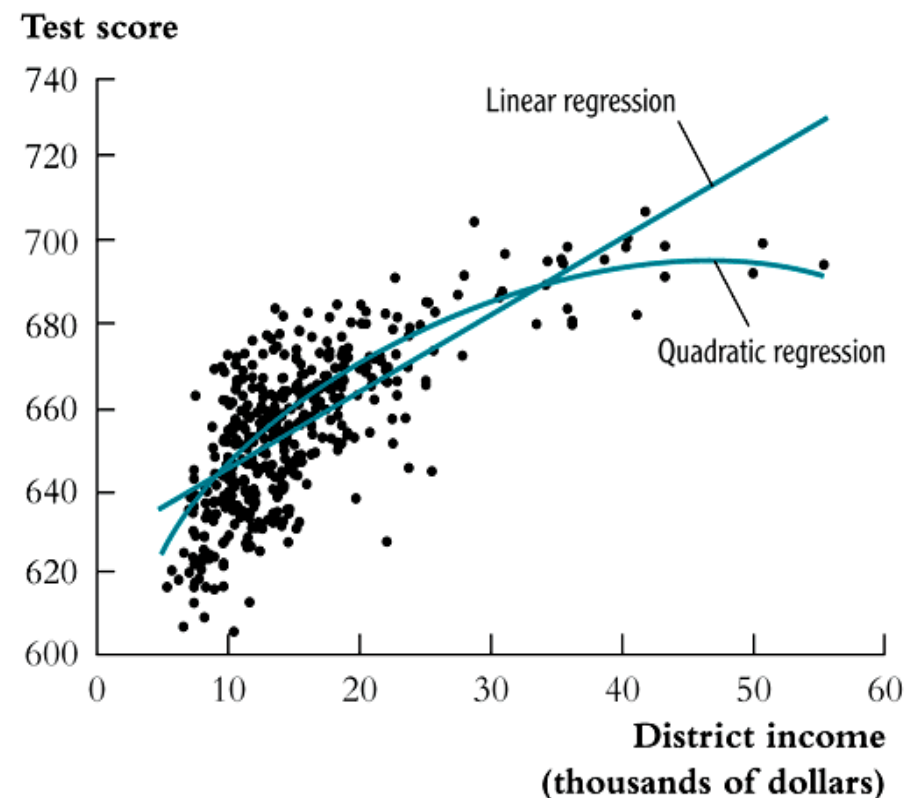
(a) Gráfico de los valores predichos

$$\widehat{TestScore} = 607.3 + 3.85Renta_i - 0.0423(Renta_i)^2$$

(2.9) (0.27) (0.0048)

FIGURE 6.3 Scatterplot of Test Score vs. District Income with Linear and Quadratic Regression Functions

The quadratic OLS regression function fits the data better than the linear OLS regression function.



Interpretación de la función de regresión estimada:

(a) Calcular los “efectos” para diferentes valores de X

$$\widehat{TestScore} = 607.3 + 3.85Renta_i - 0.0423(Renta_i)^2$$

(2.9) (0.27) (0.0048)

El cambio predicho en *Notas* para un cambio en renta de \$6,000 a \$5,000 per cápita:

$$\begin{aligned}\Delta \widehat{TestScore} &= 607.3 + 3.85 \times 6 - 0.0423 \times 6^2 \\ &\quad - (607.3 + 3.85 \times 5 - 0.0423 \times 5^2) \\ &= 3.4\end{aligned}$$

$$\widehat{TestScore} = 607.3 + 3.85Renta_i - 0.0423(Renta_i)^2$$

“Efectos” predichos para diferentes valores de X

Cambio en $Renta$ (mil\$ per cápita)	$\Delta \widehat{TestScore}$
de 5 a 6	3.4
de 25 a 26	1.7
de 45 a 46	0.0

El “efecto” de un cambio en la renta es mayor en pequeños que en niveles altos de renta (¿a lo mejor, un beneficio marginal decreciente de un incremento en los presupuestos escolares?)

Cuidado! ¿Qué ocurre con un cambio de 65 a 66?

No extrapolar fuera del rango de los datos.

Estimación de la especificación cúbica en STATA

```
gen avginc3 = avginc*avginc2;  
reg testscr avginc avginc2 avginc3, r;
```

Create the cubic regressor

Regression with robust standard errors

Number of obs = 420
F(3, 416) = 270.18
Prob > F = 0.0000
R-squared = 0.5584
Root MSE = 12.707

	Robust						
testscr	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]		

avginc	5.018677	.7073505	7.10	0.000	3.628251	6.409104	
avginc2	-.0958052	.0289537	-3.31	0.001	-.1527191	-.0388913	
avginc3	.0006855	.0003471	1.98	0.049	3.27e-06	.0013677	
_cons	600.079	5.102062	117.61	0.000	590.0499	610.108	

El término cúbico es estadísticamente significativo para el nivel de significación del 5%, pero no para el 1%

Contrastación de la hipótesis nula de linealidad frente a la hipótesis alternativa de que la regresión poblacional es cuadrática y/o cúbica, es decir, es un polinomio de grado como máximo 3:

H_0 : coeficientes poblacionales de $Renta^2$ y $Renta^3 = 0$

H_1 : al menos uno de estos coeficientes no es cero.

`test avginc2 avginc3;` **Execute the test command after running the regression**

```
( 1)  avginc2 = 0.0  
( 2)  avginc3 = 0.0
```

```
F( 2, 416) = 37.69  
Prob > F = 0.0000
```

La hipótesis de que la regresión poblacional es lineal se rechaza para el nivel de significación del 1% frente a la alternativa de que es un polinomio como máximo de grado 3.

Resumen: funciones de regresión polinomiales

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \dots + \beta_r X_i^r + u_i$$

- Estimación por MCO después de definir nuevos regresores
- Los coeficientes tienen interpretaciones complicadas
- Para interpretar la función de regresión estimada:
 - Dibujar los valores predichos como una función de x
 - Calcular los $\Delta Y/\Delta X$ predichos en diferentes valores de x
- Las hipótesis sobre el grado r pueden ser contrastadas utilizando los estadísticos t - y F - sobre la(s) variable(s) o bloques de variables apropiadas.

- Elección del grado r
 - Dibujar los datos; contrastes t - y F -, contrastar la sensibilidad de los efectos estimados; juicio.
 - *O usar los criterios de selección de modelos (después)*

2. Funciones logarítmicas de Y y/o X

- $\ln(X)$ = el logaritmo natural de X
- Las transformaciones logarítmicas permiten modelizar relaciones en términos de “porcentajes” (como elasticidades), en vez de linealmente.

¿Porqué?: $\ln(x+\Delta x) - \ln(x) = \ln\left(1 + \frac{\Delta x}{x}\right) \cong \frac{\Delta x}{x}$

(cálculo: $\frac{d \ln(x)}{dx} = \frac{1}{x}$)

Númericamente:

$$\ln(1.01) = .00995 \cong .01; \ln(1.10) = .0953 \cong .10$$

(redondeo)

Tres casos:

Caso	Función de regresión poblacional
I. lineal-log	$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$
II. log-lineal	$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$
III. log-log	$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$

- La interpretación del coeficiente de pendiente difiere en cada caso.
- La interpretación se encuentra aplicando la regla general “antes y después”: “calcular el cambio en Y para un cambio dado en X .”

I. Función de regresión poblacional lineal-log

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i \quad (b)$$

Ahora cambia X :

$$Y + \Delta Y = \beta_0 + \beta_1 \ln(X + \Delta X) \quad (a)$$

Restar (a) – (b):

$$\Delta Y = \beta_1 [\ln(X + \Delta X) - \ln(X)]$$

ahora

$$\ln(X + \Delta X) - \ln(X) \cong \frac{\Delta X}{X},$$

así

$$\Delta Y \cong \beta_1 \frac{\Delta X}{X}$$

ó

$$\beta_1 \cong \frac{\Delta Y}{\Delta X / X} \quad (\Delta X \text{ pequeño})$$

Caso lineal-log, continuación

$$Y_i = \beta_0 + \beta_1 \ln(X_i) + u_i$$

Para ΔX pequeño,

$$\beta_1 \cong \frac{\Delta Y}{\Delta X / X}$$

Ahora $100 \times \frac{\Delta X}{X}$ = cambio porcentual en X , así **un**

incremento del 1% en X (multiplicando X por 1.01) se asocia con un cambio de $.01\beta_1$ en Y .

Ejemplo: Notas vs. $\ln(Renta)$

- Primero se define el nuevo regresor, $\ln(Renta)$
- Ahora el modelo es lineal en $\ln(Renta)$, de manera que el modelo lineal-log puede estimarse por MCO:

$$\widehat{TestScore} = 557.8 + 36.42 \times \ln(Renta_i)$$

(3.8) (1.40)

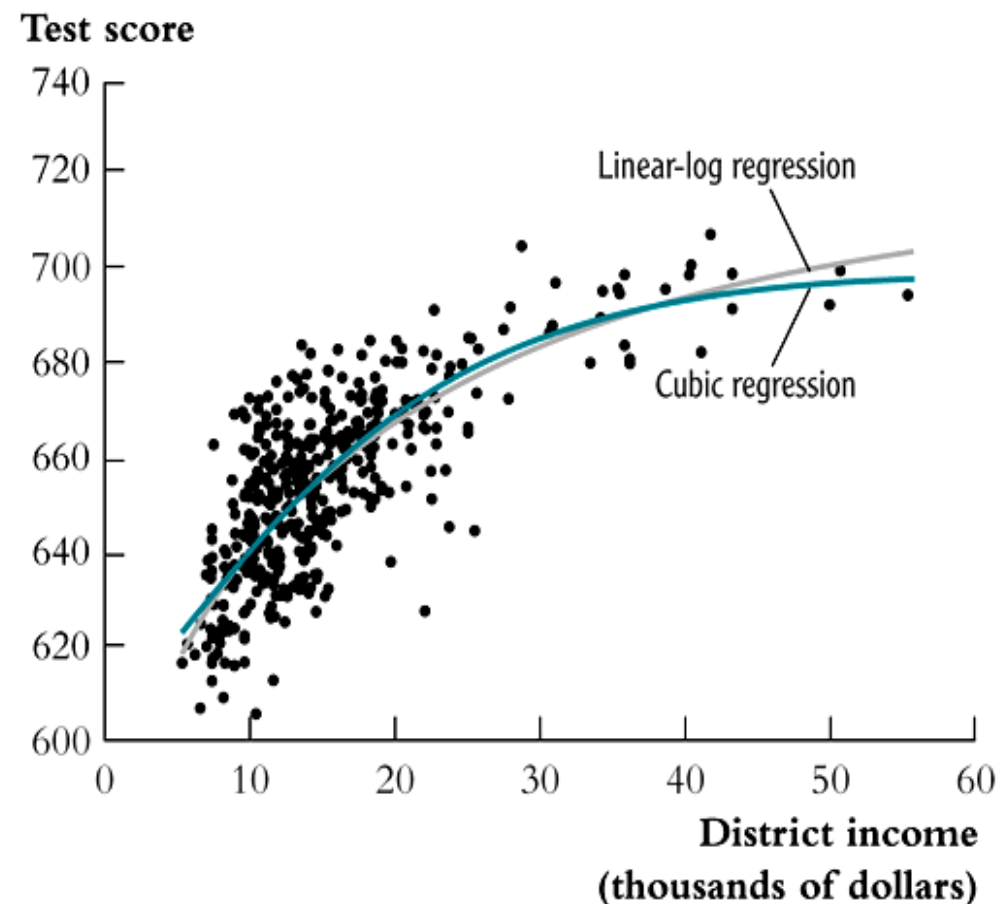
así un incremento del 1% en *Renta* se asocia con un incremento en *Notas* de 0.36 puntos en el test.

- Los errores estándar, los intervalos de confianza, R^2 – todos los instrumentos habituales de regresión se pueden aplicar aquí.
- ¿Cómo se compara este modelo con el modelo cúbico?

$$\widehat{TestScore}_i = 557.8 + 36.42 \times \ln(Renta_i)$$

FIGURE 6.7 The Linear-Log and Cubic Regression Functions

The estimated cubic regression function (Equation (6.11)) and the estimated linear-log regression function (Equation (6.18)) are nearly identical in this sample.



II. Función de regresión poblacional log-lineal

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i \quad (\text{b})$$

Ahora cambia X : $\ln(Y + \Delta Y) = \beta_0 + \beta_1(X + \Delta X) \quad (\text{a})$

Restamos (a) – (b): $\ln(Y + \Delta Y) - \ln(Y) = \beta_1 \Delta X$

así $\frac{\Delta Y}{Y} \cong \beta_1 \Delta X$

ó $\beta_1 \cong \frac{\Delta Y / Y}{\Delta X} \text{ } (\Delta X \text{ pequeño})$

caso log-lineal , continuación

$$\ln(Y_i) = \beta_0 + \beta_1 X_i + u_i$$

para ΔX pequeño, $\beta_1 \cong \frac{\Delta Y / Y}{\Delta X}$

- Ahora $100 \times \frac{\Delta Y}{Y}$ = cambio porcentual en Y , así **un cambio en X en una unidad ($\Delta X = 1$) se asocia con un cambio en Y de $100\beta_1\%$ (Y se incrementa en un factor de $1+\beta_1$).**
- *Nótese:* ¿Cuáles son las unidades de u_i y la SER?
 - Desviaciones fraccionales (proporcionales)
 - por ejemplo, $SER = .2$ significa...

III. Función de regresión poblacional log-log

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i \quad (\text{b})$$

Ahora cambia X : $\ln(Y + \Delta Y) = \beta_0 + \beta_1 \ln(X + \Delta X) \quad (\text{a})$

Restamos: $\ln(Y + \Delta Y) - \ln(Y) = \beta_1 [\ln(X + \Delta X) - \ln(X)]$

así $\frac{\Delta Y}{Y} \cong \beta_1 \frac{\Delta X}{X}$

ó $\beta_1 \cong \frac{\Delta Y / Y}{\Delta X / X} \quad (\Delta X \text{ pequeño})$

Caso log-log, continuación

$$\ln(Y_i) = \beta_0 + \beta_1 \ln(X_i) + u_i$$

para ΔX pequeño,

$$\beta_1 \cong \frac{\Delta Y / Y}{\Delta X / X}$$

Ahora $100 \times \frac{\Delta Y}{Y} =$ cambio porcentual en Y , y $100 \times \frac{\Delta X}{X} =$

cambio porcentual en X , así **un cambio del 1% en X se asocia con un cambio de $\beta_1\%$ en Y .**

- ***En la especificación log-log, β_1 tiene la interpretación de elasticidad.***

Ejemplo: $\ln(\text{Notas})$ vs. $\ln(\text{Renta})$

- Primero definimos una nueva variable dependiente, $\ln(\text{Notas})$, y el nuevo regresor, $\ln(\text{Renta})$
- El modelo es ahora una regression lineal de $\ln(\text{Notas})$ frente a $\ln(\text{Renta})$, que puede estimarse por MCO:

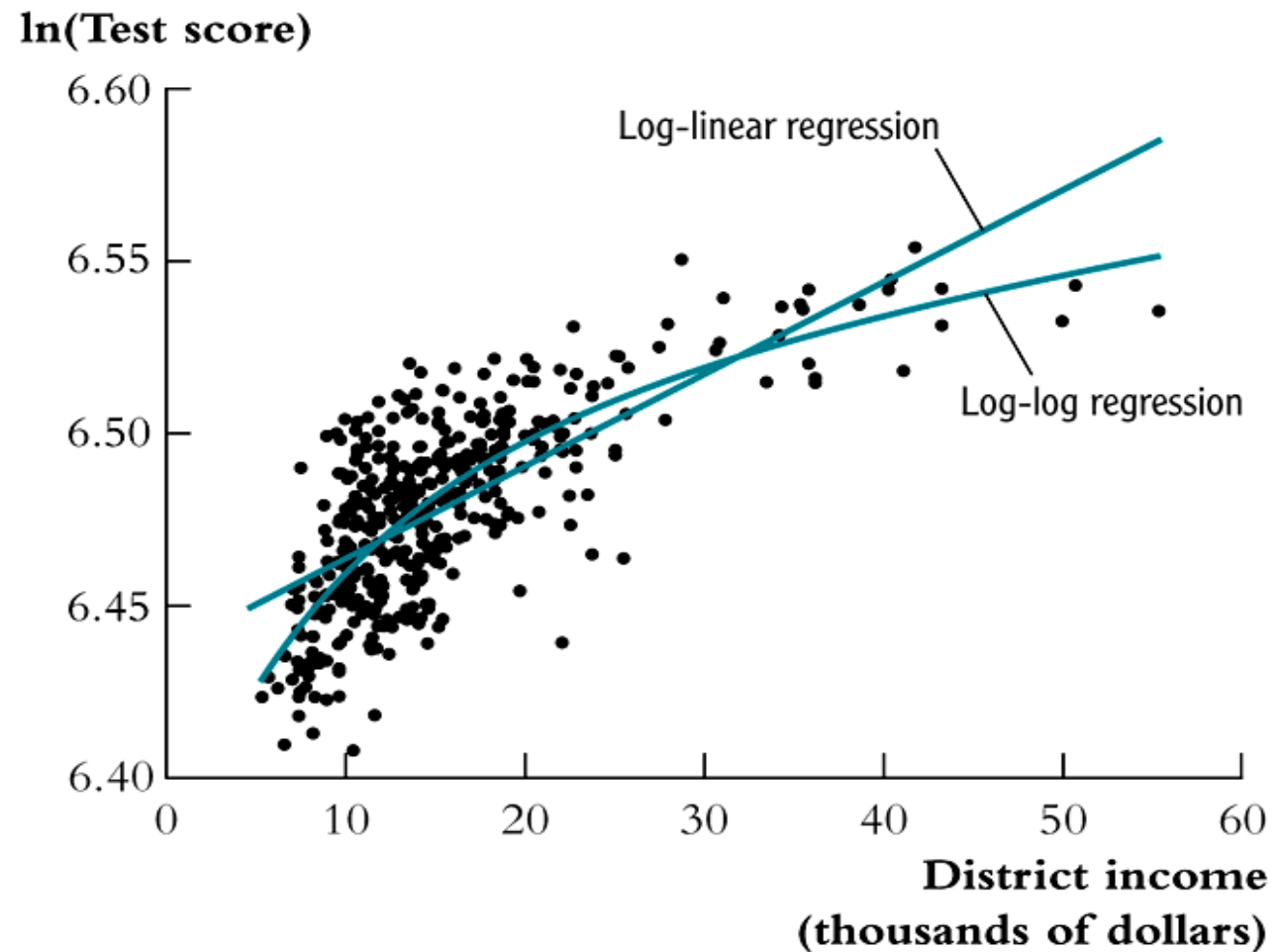
$$\widehat{\ln(\text{TestScore})} = 6.336 + 0.0554 \times \ln(\text{Renta}_i) \\ (0.006) \quad (0.0021)$$

Un incremento del 1% en *Renta* se asocia con un incremento de .0554% en *Notas* (factor de 1.0554)

- ¿Como compara este caso con el modelo log-lineal?

FIGURE 6.6 The Log-Linear and Log-Log Regression Functions

In the log-linear regression function, $\ln(Y)$ is a linear function of X . In the log-log regression function, $\ln(Y)$ is a linear function of $\ln(X)$.



Ninguna especificación parece que se ajusta tan bien como el cúbico o lineal-log

Resumen: Transformaciones logarítmicas

- Tres casos, dependiendo de si es Y y/o X es la variable transformada tomando logaritmos.
- Después de crear la(s) nueva(s) variable(s) $\ln(Y)$ y/o $\ln(X)$, la regresión es lineal en las nuevas variables y los coeficientes pueden estimarse por MCO.
- Los contrastes de hipótesis y los intervalos de confianza son ahora estándar.
- La interpretación de β_1 difiere de un caso a otro.
- La elección de la especificación debería llevarse a cabo utilizando el juicio (¿qué interpretación tiene más sentido en la aplicación?), contrastes, y graficando los valores predichos

Interacciones entre Variables Independientes

(SW Sección 6.3)

- A lo mejor una reducción del tamaño de clase es más efectiva en algunas circunstancias que en otras...
- A lo mejor clases de menor tamaño ayudan más si hay más estudiantes de inglés, que necesitan atención individual
- Es decir, $\frac{\Delta TestScore}{\Delta STR}$ puede depender de $PctEL$
- Más generalmente, $\frac{\Delta Y}{\Delta X_1}$ puede depender de X_2
- ¿Cómo modelizar estas “interacciones” entre X_1 y X_2 ?

- Consideramos primero X 's binarias, y después X 's continuas

(a) Interacciones entre dos variables binarias

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + u_i$$

- D_{1i} , D_{2i} son binarias
- β_1 es el efecto de cambiar $D_1=0$ a $D_1=1$. En esta especificación, *este efecto no depende del valor de D_2* .
- Para permitir que el efecto de cambiar D_1 dependa de D_2 , se incluye el “término de interacción” $D_{1i} \times D_{2i}$ como un regresor:

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i$$

Interpretación de los coeficientes

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 (D_{1i} \times D_{2i}) + u_i$$

Regla general: comparar los distintos casos

$$E(Y_i | D_{1i}=0, D_{2i}=d_2) = \beta_0 + \beta_2 d_2 \quad (\text{b})$$

$$E(Y_i | D_{1i}=1, D_{2i}=d_2) = \beta_0 + \beta_1 + \beta_2 d_2 + \beta_3 d_2 \quad (\text{a})$$

restar (a) – (b):

$$E(Y_i | D_{1i}=1, D_{2i}=d_2) - E(Y_i | D_{1i}=0, D_{2i}=d_2) = \beta_1 + \beta_3 d_2$$

- El efecto de D_1 depende de d_2 (lo que queríamos)
- β_3 = incremento del efecto de D_1 , cuando $D_2 = 1$

Ejemplo: Notas, STR, Estudiantes de Inglés

Sea

$$STR = \begin{cases} 1 & \text{si } STR \geq 20 \\ 0 & \text{si } STR < 20 \end{cases} \quad \text{y} \quad HiEL = \begin{cases} 1 & \text{si } PctEL \geq 10 \\ 0 & \text{si } PctEL < 10 \end{cases}$$

$$\widehat{TestScore} = 664.1 - 18.2HiEL - 1.9HiSTR - 3.5(HiSTR \times HiEL)$$

(1.4) (2.3) (1.9) (3.1)

- “Efecto” of *HiSTR* cuando *HiEL* = 0 es -1.9
- “Efecto” of *HiSTR* cuando *HiEL* = 1 es $-1.9 - 3.5 = -5.4$

- La reducción del tamaño de clase se estima para tener un mayor efecto cuando el porcentaje de estudiantes de inglés es grande
- Esta interacción no es estadísticamente significativa: $t = 3.5/3.1$

(b) Interacciones entre variables continuas y binarias

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + u_i$$

- D_i es binaria, X es continua
- Tal y como se especificó antes, el efecto en Y de X (manteniendo D constante) $= \beta_2$, el cual no depende de D

- Para permitir que el efecto de X dependa de D , se incluye el “término de interacción” $D_i \times X_i$ como regresor:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i \times X_i) + u_i$$

Interpretación de los coeficientes

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i \times X_i) + u_i$$

Regla general: comparar varios casos

$$Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 (D \times X) \quad (b)$$

Ahora cambia X :

$$Y + \Delta Y = \beta_0 + \beta_1 D + \beta_2 (X + \Delta X) + \beta_3 [D \times (X + \Delta X)] \quad (a)$$

restar (a) – (b):

$$\Delta Y = \beta_2 \Delta X + \beta_3 D \Delta X \quad \text{ó} \quad \frac{\Delta Y}{\Delta X} = \beta_2 + \beta_3 D$$

- El efecto de X depende de D (lo que queríamos)
- β_3 = incremento del efecto de X , cuando $D = 1$

Ejemplo: Notas, STR, HiEL (=1 si PctEL ≥ 20)

$$\widehat{TestScore} = 682.2 - 0.97STR + 5.6HiEL - 1.28(STR \times HiEL)$$

(11.9) (0.59) (19.5) (0.97)

- Cuando $HiEL = 0$:

$$\widehat{TestScore} = 682.2 - 0.97STR$$

- Cuando $HiEL = 1$,

$$\begin{aligned}\widehat{TestScore} &= 682.2 - 0.97STR + 5.6 - 1.28STR \\ &= 687.8 - 2.25STR\end{aligned}$$

- Dos líneas de regresión: una para cada grupo $HiSTR$.
- La reducción del tamaño de clase se estima que tiene un mayor efecto cuando el porcentaje de estudiantes de inglés es mayor.

Ejemplo, continuación.

$$\widehat{TestScore} = 682.2 - 0.97STR + 5.6HiEL - 1.28(STR \times HiEL)$$

(11.9) (0.59) (19.5) (0.97)

Contraste de varias hipótesis:

- Las dos líneas de regresión tienen la misma pendiente
 \Leftrightarrow el coeficiente de $STR \times HiEL$ es cero:

$$t = -1.28/0.97 = -1.32 \Rightarrow \text{no se puede rechazar}$$

- Las dos líneas de regresión tienen el mismo término constante \Leftrightarrow el coeficiente de $HiEL$ es cero:

$$t = -5.6/19.5 = 0.29 \Rightarrow \text{no se puede rechazar}$$

Ejemplo, continuación.

$$\widehat{TestScore} = 682.2 - 0.97STR + 5.6HiEL - 1.28(STR \times HiEL),$$

(11.9) (0.59) (19.5) (0.97)

- **Hipótesis conjunta** de que las dos líneas de regresión son las mismas \Leftrightarrow coeficiente poblacional de $HiEL = 0$ y el coeficiente poblacional de $STR \times HiEL = 0$:

$$F = 89.94 \text{ (} p\text{-value} < .001 \text{) !!}$$

- ¿Por qué rechazamos la hipótesis conjunta pero no las hipótesis individuales?
- Consecuencia de alta pero imperfecta multicolinealidad: correlación alta entre $HiEL$ y $STR \times HiEL$

Interacciones binaria-continua: las dos líneas de regresión

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_i + \beta_3 (D_i \times X_i) + u_i$$

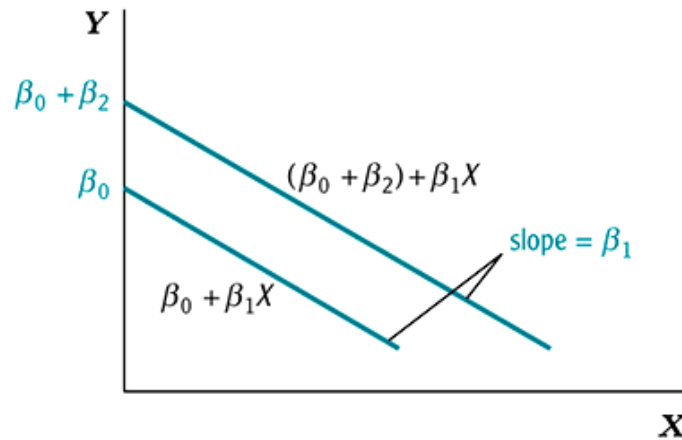
Observaciones con $D_i = 0$ (el grupo “ $D = 0$ ”):

$$Y_i = \beta_0 + \beta_2 X_i + u_i$$

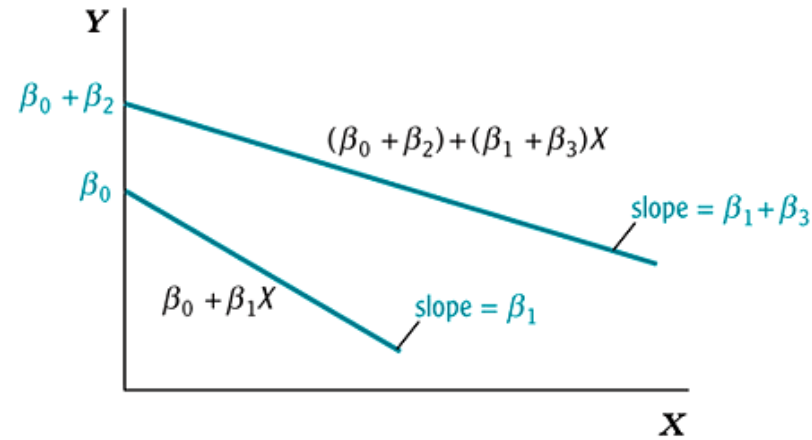
Observaciones con $D_i = 1$ (el grupo “ $D = 1$ ”):

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 + \beta_2 X_i + \beta_3 X_i + u_i \\ &= (\beta_0 + \beta_1) + (\beta_2 + \beta_3) X_i + u_i \end{aligned}$$

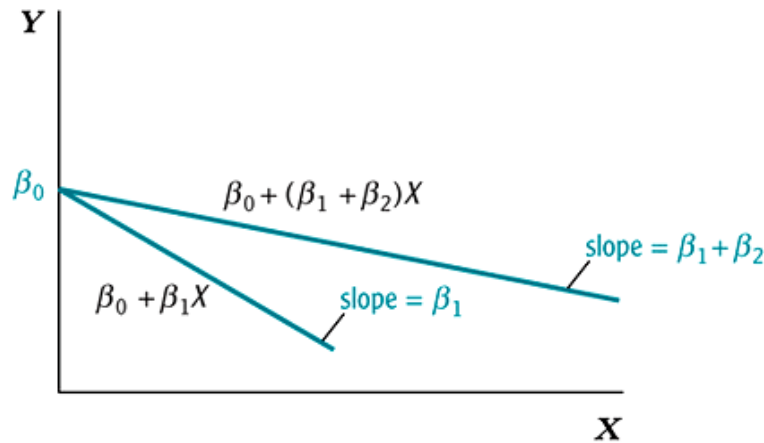
FIGURE 6.8 Regression Functions Using Binary and Continuous Variables



(a) Different intercepts, same slope



(b) Different intercepts, different slopes



(c) Same intercept, different slopes

Interactions of binary variables and continuous variables can produce three different population regression functions: (a) $\beta_0 + \beta_1 X + \beta_2 D$ allows for different intercepts but has the same slope; (b) $\beta_0 + \beta_1 X + \beta_2 D + \beta_3 (X \times D)$ allows for different intercepts and different slopes; and (c) $\beta_0 + \beta_1 X + \beta_2 (X \times D)$ has the same intercept but allows for different slopes.

(c) Interacciones entre dos variables continuas

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

- X_1, X_2 son continuas
- Como se especificó, el efecto de X_1 no depende de X_2
- Como se especificó, el efecto de X_2 no depende de X_1
- Para permitir que el efecto de X_1 dependa de X_2 , se incluye el “término de interacción” $X_{1i} \times X_{2i}$ como regresor:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i$$

Coeficientes en interacciones continua-continua

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 (X_{1i} \times X_{2i}) + u_i$$

Regla general: comparar los diferentes casos

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 \times X_2) \quad (b)$$

Ahora cambia X_1 :

$$Y + \Delta Y = \beta_0 + \beta_1 (X_1 + \Delta X_1) + \beta_2 X_2 + \beta_3 [(X_1 + \Delta X_1) \times X_2] \quad (a)$$

restar (a) – (b):

$$\Delta Y = \beta_1 \Delta X_1 + \beta_3 X_2 \Delta X_1 \quad \text{ó} \quad \frac{\Delta Y}{\Delta X_1} = \beta_2 + \beta_3 X_2$$

- El efecto de X_1 depende de X_2 (lo que queríamos)
- β_3 = incremento del efecto de X_1 a partir de un cambio unitario en X_2

Ejemplo: Notas, STR, PctEL

$$\widehat{TestScore} = 686.3 - 1.12STR - 0.67PctEL + .0012(STR \times PctEL),$$

(11.8) (0.59) (0.37) (0.019)

El efecto estimado de la reducción del tamaño de clase es no lineal porque el tamaño del efecto depende de *PctEL*:

$$\frac{\Delta TestScore}{\Delta STR} = -1.12 + .0012PctEL$$

<i>PctEL</i>	$\frac{\Delta TestScore}{\Delta STR}$
0	-1.12
20%	$-1.12 + .0012 \times 20 = -1.10$

Ejemplo, continuación: contrastes de hipótesis

$$\widehat{TestScore} = 686.3 - 1.12STR - 0.67PctEL + .0012(STR \times PctEL),$$

(11.8) (0.59) (0.37) (0.019)

- ¿Es el coeficiente poblacional de $STR \times PctEL = 0$?

$$t = .0012/.019 = .06 \Rightarrow \text{no se puede rechazar la nula al 5\%}.$$

- ¿Es el coeficiente poblacional de $STR = 0$?

$$t = -1.12/0.59 = -1.90 \Rightarrow \text{no se puede rechazar la nula al 5\%}$$

- ¿Son los coeficientes ***tanto en STR como en STR × PctEL*** = 0?

$F = 3.89$ (p -value = .021) \Rightarrow se rechaza la nula al nivel del 5%(!!) (¿Por qué? Alta pero imperfecta multicolinealidad)

Aplicación: Efectos No Lineales en las Notas Del Ratio Alumno-Profesor (SW Sección 6.4)

Nos centramos en dos cuestiones:

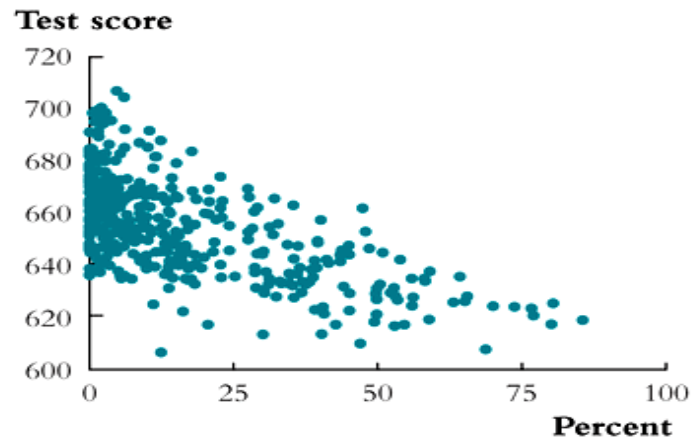
1. ¿Hay efectos no lineales de la reducción del tamaño de la clase sobre los resultados del test? (¿Tiene una reducción de 35 a 30 el mismo efecto que una reducción de 20 a 15?)
2. ¿Hay interacciones no lineales entre *PctEL* y *STR*? (¿Son las clases pequeñas más efectivas cuando hay muchos estudiantes de Inglés?)

Estrategia para la Cuestión #1 (diferentes efectos para diferente *STR*?)

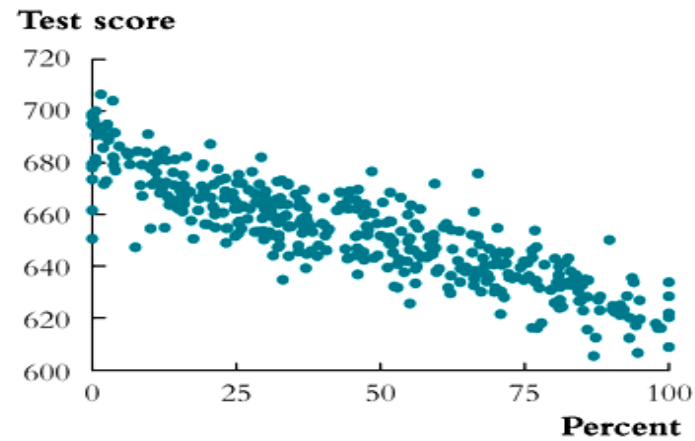
- Estimar las funciones lineal y no lineal de *STR*,
manteniendo constantes las variables demográficas
relevantes
 - *PctEL*
 - *Renta* (recordar la relación no lineal *Notas-renta*!)
 - *LunchPCT* (parte de comida gratis/subvencionada)
- Ver si al añadir los términos no lineales se genera una
diferencia cuantitativa “económicamente importante”
(importancia “económica” ó “del mundo-real” es diferente
de la significatividad estadística)
- Contrastar si los términos no lineales son significativos

¿Qué es una buena especificación “base”?

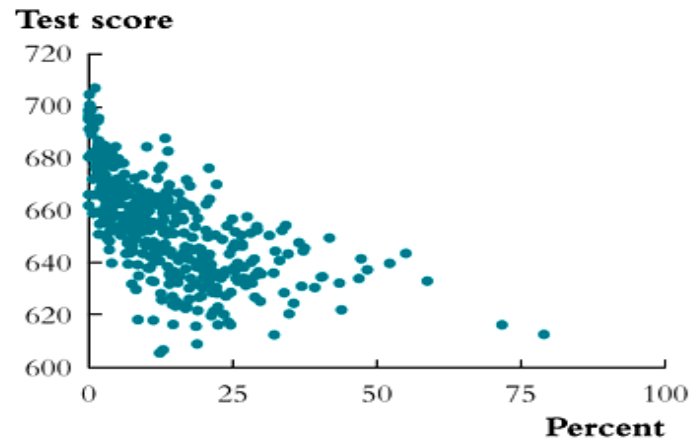
FIGURE 5.2 Scatterplots of Test Scores vs. Three Student Characteristics



(a) Percent of English language learners



(b) Percent qualifying for reduced price lunch



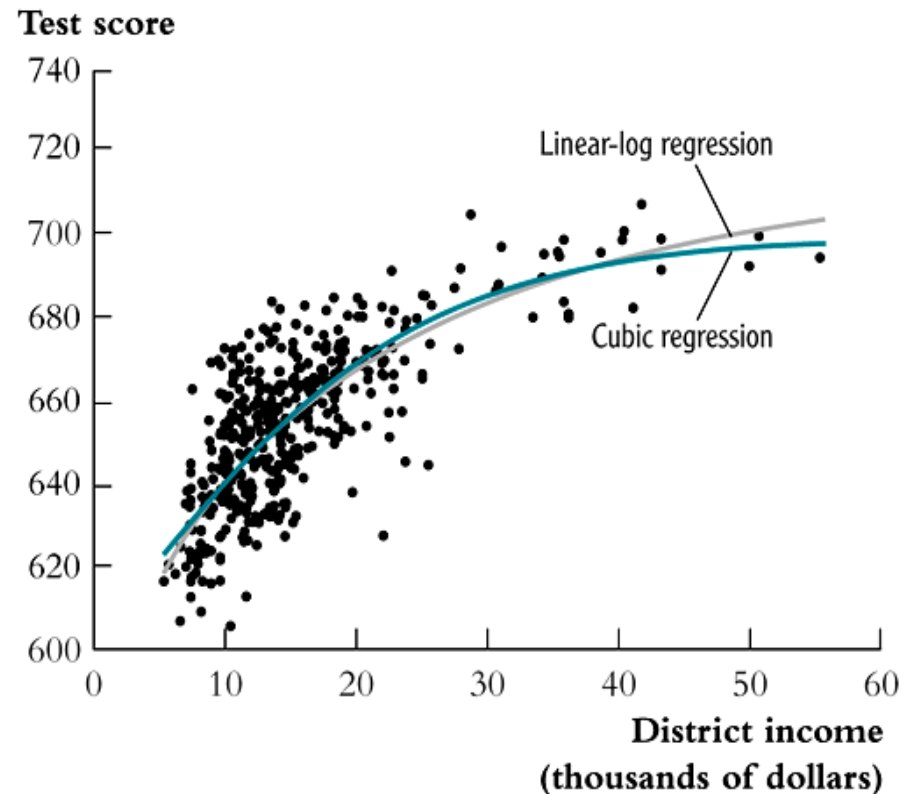
(c) Percent qualifying for income assistance

The scatterplots show a negative relationship between test scores and (a) the percentage of English learners (correlation = -0.64), (b) the percentage of students qualifying for a subsidized lunch (correlation = -0.87); and (c) the percentage qualifying for income assistance (correlation = -0.63).

La relación *Notas – Renta*

FIGURE 6.7 The Linear-Log and Cubic Regression Functions

The estimated cubic regression function (Equation (6.11)) and the estimated linear-log regression function (Equation (6.18)) are nearly identical in this sample.



Una ventaja de la especificación logarítmica es que se comporta mejor cerca del final de la muestra, especialmente para valores grandes de la renta.

Especificación Base

A partir de los diagramas de puntos y del análisis anterior, aquí hay posibles puntos de partida para las variables de control demográfico:

Variable dependiente: *Notas*

Variable Independiente	Forma Funcional
<i>PctEL</i>	Lineal
<i>LunchPCT</i>	Lineal
<i>Renta</i>	ln(Renta) (ó se podría usar cúbica)

Cuestión #1:

Investigar considerando un polinomio en STR

$$\begin{aligned}\widehat{TestScore} = & 252.0 + 64.33STR - 3.42STR^2 + .059STR^3 \\ & (163.6) \quad (24.86) \quad (1.25) \quad (.021) \\ & - 5.47HiEL - .420LunchPCT + 11.75\ln(Renta) \\ & (1.03) \quad (.029) \quad (1.78)\end{aligned}$$

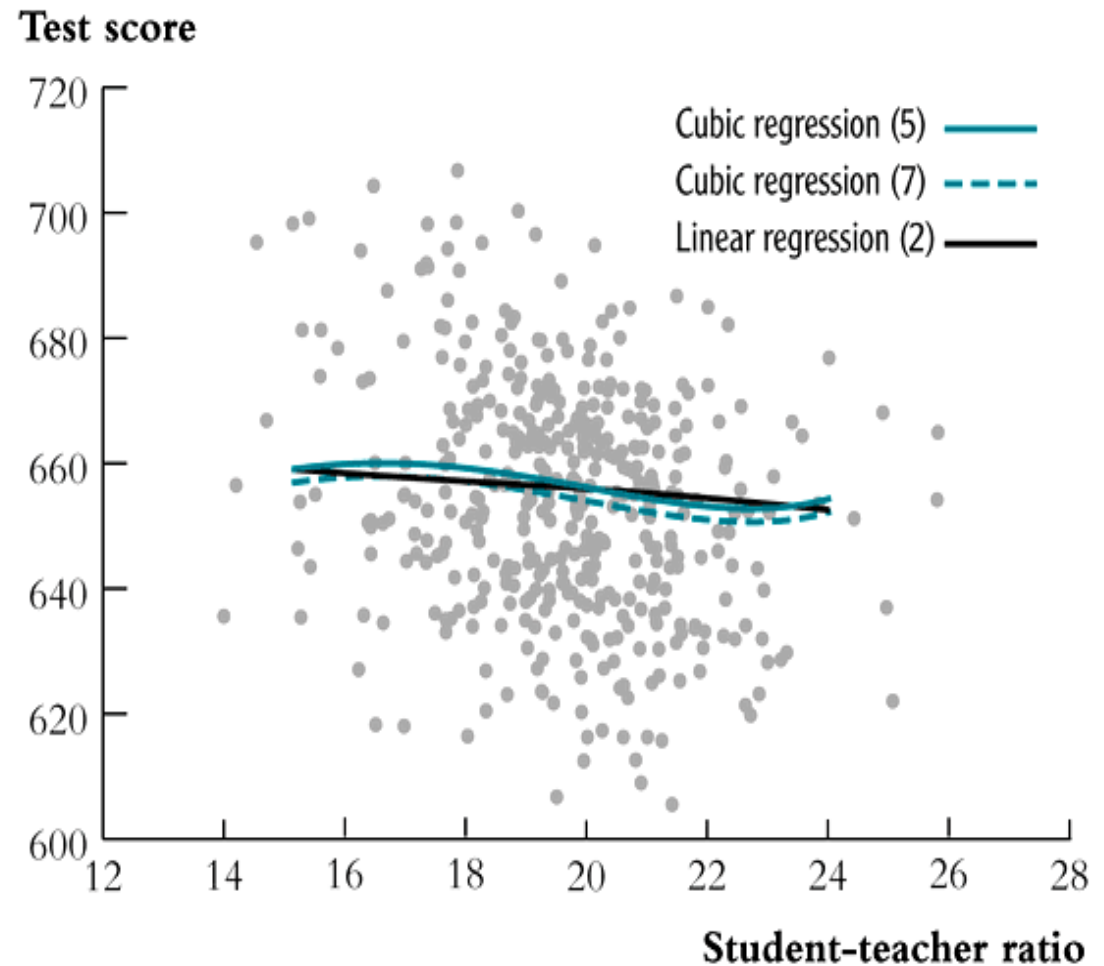
Interpretación de los coeficientes de:

- $HiEL$?
- $LunchPCT$?
- $\ln(Renta)$?
- STR, STR^2, STR^3 ?

Interpretar la función de regresión vía gráficos (la regresión anterior se denota por (5) en esta figura)

FIGURE 6.10 Three Regression Functions Relating Test Scores and Student-Teacher Ratio

The cubic regressions from columns (5) and (7) of Table 6.2 are nearly identical. They indicate a small amount of nonlinearity in the relation between test scores and student-teacher ratio.



¿Son los términos de mayor orden en *STR* estadísticamente significativos?

$$\begin{aligned}\widehat{TestScore} = & 252.0 + 64.33STR - 3.42STR^2 + .059STR^3 \\ & (163.6) \quad (24.86) \quad (1.25) \quad (.021) \\ & - 5.47HiEL - .420LunchPCT + 11.75\ln(Renta) \\ & (1.03) \quad (.029) \quad (1.78)\end{aligned}$$

(a) H_0 : cuadrático en *STR* frente a H_1 : cúbico en *STR*?

$$t = .059/.021 = 2.86 \quad (p = .005)$$

(b) H_0 : lineal en *STR* frente a H_1 : no lineal/hasta el cúbico en *STR*?

$$F = 6.17 \quad (p = .002)$$

Cuestión #2: interacciones *STR-PctEL*

(para simplificar las cosas, se ignoran los términos STR^2 , STR^3 por ahora)

$$\begin{aligned}\widehat{TestScore} = & 653.6 - .53STR + 5.50HiEL - .58HiEL \times STR \\ & (9.9) \quad (.34) \quad (9.80) \quad (.50) \\ & - .411LunchPCT + 12.12\ln(Renta) \\ & (.029) \quad (1.80)\end{aligned}$$

Interpretación de los coeficientes en:

- *STR*?
- *HiEL*? (¿signo incorrecto?)
- *HiEL* × *STR*?
- *LunchPCT*?
- $\ln(Renta)$?

Interpretación de las funciones de regresión vía gráficos:

$$\widehat{TestScore} = 653.6 - .53STR + 5.50HiEL - .58HiEL \times STR$$

(9.9) (.34) (9.80) (.50)

$$- .411LunchPCT + 12.12\ln(Renta)$$

(.029) (1.80)

Importancia en el “Mundo-Real” (“política” ó “económica”) del término de interacción:

$$\frac{\Delta \widehat{TestScore}}{\Delta STR} = -.53 - .58HiEL = \begin{cases} -1.12 & \text{si } HiEL = 1 \\ -.53 & \text{si } HiEL = 0 \end{cases}$$

- La diferencia en el efecto estimado de reducir el *STR* es sustancial; la reducción en el tamaño de clase es más efectiva en distritos con más estudiantes de inglés

¿El término del efecto de interacción es estadísticamente significativo?

$$\begin{aligned}\widehat{TestScore} = & 653.6 - .53STR + 5.50HiEL - .58HiEL \times STR \\ & (9.9) \quad (.34) \quad (9.80) \quad (.50) \\ & - .411LunchPCT + 12.12\ln(Renta) \\ & (.029) \quad (1.80)\end{aligned}$$

(a) H_0 : coef. de interacción=0 frente a H_1 : interacción no cero

$$t = -1.17 \Rightarrow \text{no significativo al 10\%}$$

(b) H_0 : ambos coefs asociados con $STR = 0$ frente a H_1 : al menos un coef. es distinto de 0 (STR entra)

$$F = 5.92 (p = .003)$$

Siguiente: especificaciones con polinomios + interacciones!

TABLE 6.2 Nonlinear Regression Models of Test Scores**Dependent Variable: Average Test Score in District; 420 Observations.**

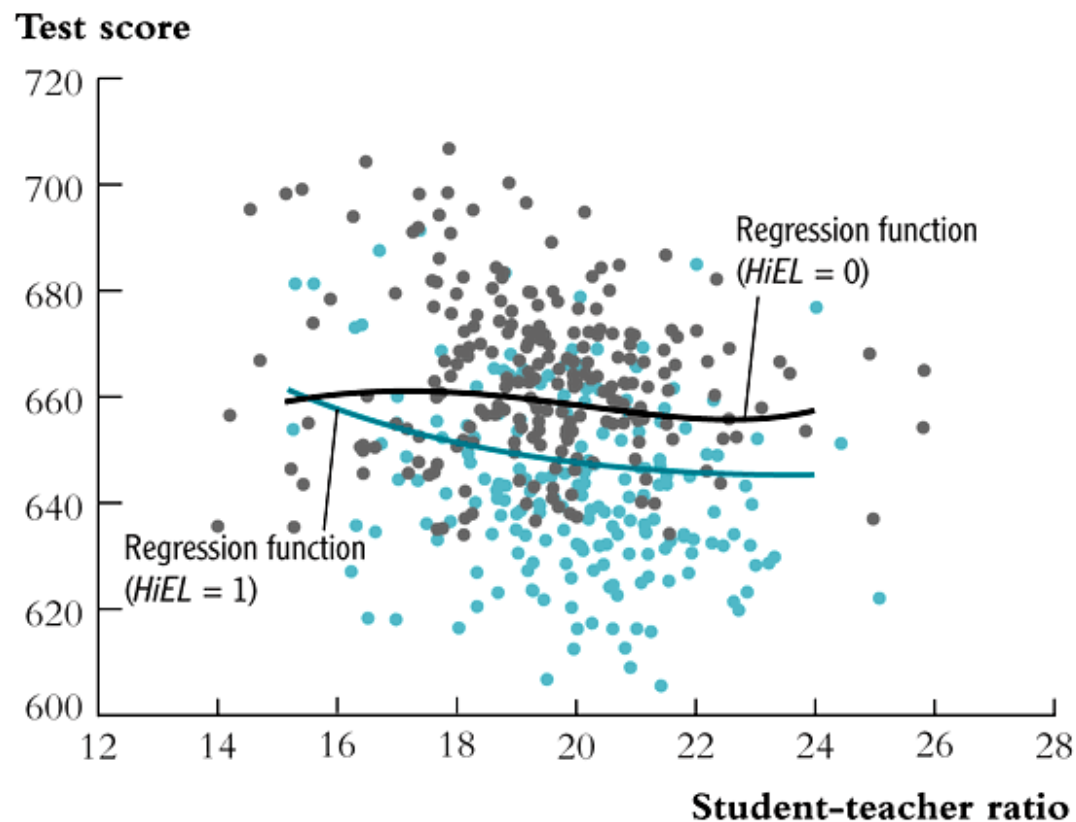
Regressor	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Student-teacher ratio (<i>STR</i>)	-1.00** (0.27)	-0.73** (0.26)	-0.97 (0.59)	-0.53 (0.34)	64.33** (24.86)	83.70** (28.50)	65.29** (25.26)
STR^2					-3.42** (1.25)	-4.38** (1.44)	-3.47** (1.27)
STR^3					0.059** (0.021)	0.075** (0.024)	0.060** (0.021)
% English Learners	-0.122** (0.033)	-0.176** (0.034)					-0.166** (0.034)
% English Learners ≥10%? (Binary, <i>HiEL</i>)			5.64 (19.51)	5.50 (9.80)	-5.47** (1.03)	816.1* (327.7)	
$HiEL \times STR$			-1.28 (0.97)	-0.58 (0.50)		-123.3* (50.2)	
$HiEL \times STR^2$						6.12* (2.54)	
$HiEL \times STR^3$						-0.101* (0.043)	
% Eligible for subsidized lunch	-0.547** (0.024)	-0.398** (0.033)		-0.411** (0.029)	-0.420** (0.029)	-0.418** (0.029)	-0.402** (0.033)
Average district income (logarithm)		11.57** (1.81)		12.12** (1.80)	11.75** (1.78)	11.80** (1.78)	11.51** (1.81)
Intercept	700.1** (5.6)	658.6** (8.6)	682.2** (11.9)	653.6** (9.9)	252.0 (163.6)	122.3 (185.5)	244.8 (165.7)

These regressions were estimated using the data on K–8 school districts in California, described in Appendix 4.1. Standard errors are given in parentheses under coefficients, and *p*-values are given in parentheses under *F*-statistics. Individual coefficients are statistically significant at the *5% or **1% significance level.

Interpretación de las funciones de regresión vía gráficos

FIGURE 6.11 Regression Functions for Districts with High and Low Percentages of English Learners

Districts with low percentages of English learners ($HiEL = 0$) are shown by gray dots and districts with $HiEL = 1$ are shown by colored dots. The cubic regression function for $HiEL = 1$ from regression (6) in Table 6.2 is approximately 10 points below the cubic regression function for $HiEL = 0$ for $17 \leq STR \leq 23$, but otherwise the two functions have similar shapes and slopes in this range. The slopes of the regression functions differ most for very large and small values of STR , where there are few observations.



Contrastes de hipótesis conjuntas:

TABLE 6.2 Nonlinear Regression Models of Test Scores

Dependent Variable: Average Test Score in District; 420 Observations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
F-statistics and p-values on Joint Hypotheses							
(a) all <i>STR</i> variables and interactions = 0			5.64 (0.004)	5.92 (0.003)	6.31 (<0.001)	4.96 (<0.001)	5.91 (0.001)
(b) $STR^2, STR^3 = 0$					6.17 (<0.001)	5.81 (0.003)	5.96 (0.003)
(c) $HiEL \times STR, HiEL \times STR^2,$ $HiEL \times STR^3 = 0$						2.69 (0.046)	
<i>SER</i>	9.08	8.64	15.88	8.63	8.56	8.55	8.57
\overline{R}^2	0.773	0.794	0.305	0.795	0.798	0.799	0.798

These regressions were estimated using the data on K–8 school districts in California, described in Appendix 4.1. Standard errors are given in parentheses under coefficients, and *p*-values are given in parentheses under *F*-statistics. Individual coefficients are statistically significant at the *5% or **1% significance level.

Resumen: Funciones de Regresión No Lineales

- El uso de funciones de las variables independientes tales como $\ln(X)$ ó $X_1 \times X_2$, permite reformular una extensa familia de funciones de regresión no lineales como la regresión múltiple.
- La estimación y la inferencia se realizan de la misma manera que en modelo de regresión lineal múltiple.
- La interpretación de los coeficientes es específica en cada modelo, pero la regla general es calcular los efectos comparando diferentes casos (valores diferentes de las X 's originales)

- Muchas especificaciones no lineales son posibles, de manera que se debe utilizar el juicio: ¿Qué efecto no lineal deseamos analizar? ¿Qué sentido tiene en la aplicación en concreto?