

Clase auxiliar # 10

Regresión discontinua nítida y difusa

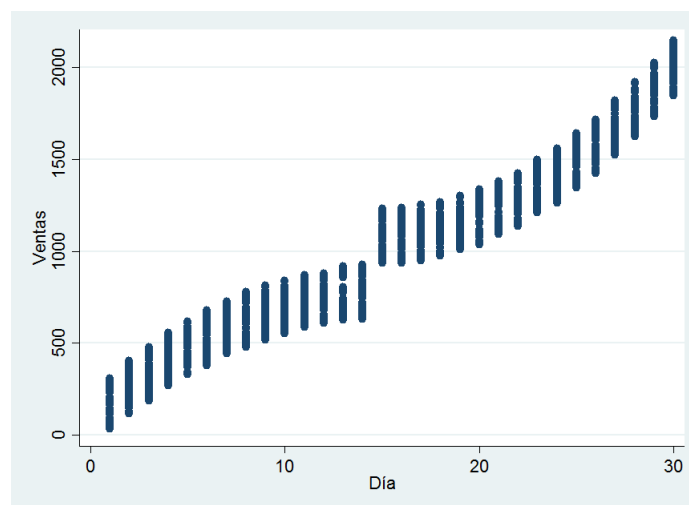
1. Regresión discontinua (1 discontinuidad)

El Gerente de Ventas de una famosa firma de juguetes quiere conocer el comportamiento de la demanda por *spinners* durante el mes de diciembre, de cara a la navidad. La empresa en cuestión tiene 50 multitiendas –convenientemente todas iguales– distribuidas a lo largo del país. Las ventas de cada una de estas empresas entre los días 1 y 25 de diciembre se encuentran registradas en la base de datos Auxiliar 10 Regresión Discontinua BD.dta.

A modo de antecedente, dicho gerente le comenta que el día 15 de diciembre se llevó a cabo una espectacular campaña de Marketing, y sospecha que esta misma pudo tener un efecto en la cantidad de compras efectuadas.

- Obtenga un gráfico de dispersión (o *scatter*) que le permita visualizar cómo se comportó la demanda entre los días 1 y 25 de diciembre.

Respuesta: Ver archivo Auxiliar 10 Código.do.



- Proponga un modelo de regresión discontinua para dicha demanda. Explique a través de esperanzas condicionales la estructura del modelo.

Respuesta: sea Y_n la cantidad de ventas registradas en la observación n , X_n el día al cual corresponde la observación n (entre 1 y 25) y $D_n = \mathbb{I}_{X_n \geq 15}$. A continuación se presenta un posible modelo para explicar el comportamiento de la demanda por *spinners* y los valores esperados de las ventas según el mismo modelo en cada tramo.

$$Y_n = \beta_0 + \tau D_n + \beta_1(X_n - 15)(1 - D_n) + \beta_2(X_n - 15)^2(1 - D_n) + \beta_3(X_n - 15)D_n + \beta_4(X_n - 15)^2D_n + U_n$$

$$\mathbb{E}[Y_n|X_n < 15] = \beta_0 + \beta_1(X_n - 15) + \beta_2(X_n - 15)^2$$

$$\mathbb{E}[Y_n|X_n \geq 15] = \beta_0 + \beta_3(X_n - 15) + \beta_4(X_n - 15)^2$$

A través de las esperanzas condicionales se observa que la estructura de este modelo consiste en 2 funciones cuadráticas distintas para cada tramo. Observando el gráfico de dispersión de la parte anterior, se puede intuir que el modelo –una vez estimado– será creciente y cóncavo en el primer tramo y creciente y convexo en el segundo tramo.

- c. Ejecute este modelo de regresión en *Stata* y estime el efecto de la propaganda sobre la demanda $\hat{\tau}$. ¿Es significativo? Luego sobreponga la predicción al gráfico de dispersión (*scatter*). Nota: puede ocupar como variable independiente tanto los días de venta como la diferencia entre estos y el día 15, en el que se produce la discontinuidad.

Respuesta:

Ocupando día como variable independiente

Ver archivo Auxiliar 10 Código.do. A partir de los datos del cuadro 3, se puede calcular el efecto inmediato de la propaganda sobre la demanda (τ)¹.

	(1)	(2)
	Ventas	
VARIABLES	coef	se
Propaganda = 1	1,883***	(163.8)
0b.propaganda#c.dia	93.41***	(3.471)
1.propaganda#c.dia	-118.0***	(16.63)
0b.propaganda#c.dia_2	-3.163***	(0.225)
1.propaganda#c.dia_2	3.951***	(0.415)
Constant	89.31***	(11.32)
Observations	1,250	
R-squared	0.950	

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

$$\hat{\tau} \approx \underbrace{(89,31 + 1883 - 118 \cdot 15 + 3,951 \cdot 15^2)}_{\lim_{d \rightarrow 15+} \hat{\mathbb{E}}[ventas|dia=d]} - \underbrace{(89,31 + 93,41 \cdot 15 - 3,163 \cdot 15^2)}_{\lim_{d \rightarrow 15-} \hat{\mathbb{E}}[ventas|dia=d]} \approx 312,4$$

¹Se usó en lugar de = pues se reemplazaron los valores extraídos del cuadro 3, los cuales están truncados y redondeados por *Stata*. Esto implica leves imprecisiones.

Como se puede ver en la imagen 1, la discontinuidad dada por $\hat{\tau}$ es estadísticamente significativa al 95 % de confianza.

```
. test _cons + 1.propaganda + 15 * 1.propaganda#dia + 225 * 1.propaganda#dia_2 == _cons + 15 * 0.propaganda#dia
> + 225 * 0.propaganda#dia_2

( 1)  1.propaganda - 15*0b.propaganda#c.dia + 15*1.propaganda#c.dia - 225*0b.propaganda#c.dia_2 +
      225*1.propaganda#c.dia_2 = 0

      F( 1, 1244) = 456.65
      Prob > F = 0.0000
```

Figura 1: test de restricción lineal para evaluar si la discontinuidad dada por $\hat{\tau}$ es estadísticamente significativa al 95 % de confianza.

Notar que los términos cuadráticos están acompañados de un coeficiente negativo ($-3,163$) antes de la discontinuidad y positivo ($3,951$) después de ella. Esto refleja la forma cóncava/convexa en la que se comporta la muestra antes y después de la discontinuidad (ver figura 1 de la parte a).

La superposición del modelo de regresión discontinua con gráfico de dispersión se encuentra en la figura 3.

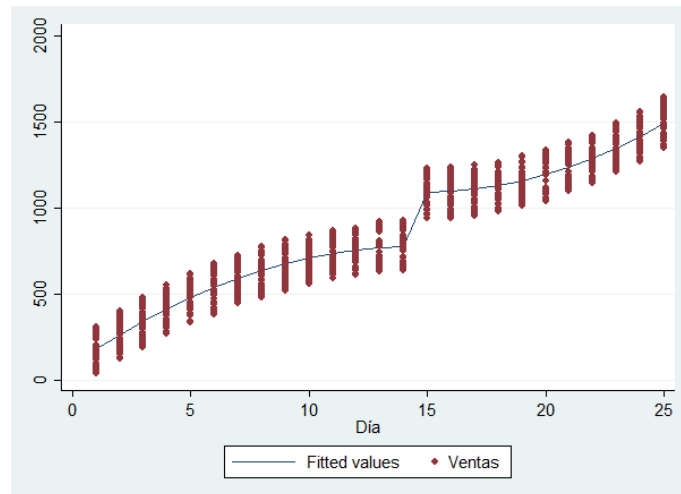


Figura 2: superposición del modelo de regresión discontinua con gráfico de dispersión. La variable independiente es el día.

Ocupando (día – 15) como variable independiente

Ver archivo Auxiliar 10 Código.do. Como se puede observar en el cuadro 1, el efecto inmediato de la propaganda sobre la demanda (τ) se estima en $\hat{\tau} = 312,4$ unidades de *spinners* vendidos y es estadísticamente significativo con un 95 % de confianza.

Notar que los términos cuadráticos están acompañados de un coeficiente negativo ($-3,163$) antes de la discontinuidad y positivo ($3,951$) después de ella. Esto refleja la forma cóncava/convexa en la que se comporta la muestra antes y después de la discontinuidad (ver figura 1 de la parte a).

VARIABLES	(1) Ventas coef	(2) se
Propaganda = 1	312.4***	(14.62)
0b.propaganda#c.dif_dias	-1.495	(3.471)
1.propaganda#c.dif_dias	0.522	(4.306)
0b.propaganda#c.dif_dias_2	-3.163***	(0.225)
1.propaganda#c.dif_dias_2	3.951***	(0.415)
Constant	778.6***	(11.32)
Observations	1,250	
R-squared	0.950	

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

La superposición del modelo de regresión discontinua con gráfico de dispersión se encuentra en la figura 1.

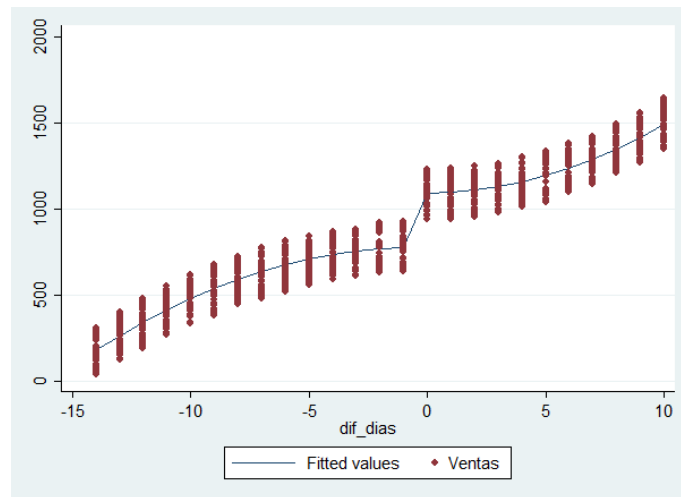


Figura 3: superposición del modelo de regresión discontinua con gráfico de dispersión. La variable independiente es el día.

- d. Ocupando el comando `rd`, realice esta vez una regresión discontinua local polinómica y estime el efecto de la propaganda sobre la demanda $\hat{\tau}$. ¿Es significativo? Luego sobreponga la predicción al gráfico de dispersión (*scatter*). Nota: puede ocupar como variable independiente tanto los días de venta como la diferencia entre estos y el día 15, en el que se produce la discontinuidad.

Respuesta: esto se puede hacer de dos formas: ocupando la variable independiente con² y sin desplazar³. En el primer caso, hay que declarar que el punto de discontinuidad es 0. En el segundo caso, hay que declarar que el punto de discontinuidad es 15. Los estadísticos y el gráfico resultante son los mismos en ambos casos. Como se puede observar en la figura 4, el efecto estimado de la propaganda en las ventas es de $\hat{\tau} \approx 309,14$ unidades, el cual es estadísticamente significativo.

Estimating for bandwidth 3.49593627961054

ventas	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
lwald	309.1395	23.67397	13.06	0.000	262.7394	355.5396

Figura 4: estimación $\hat{\tau}$ del efecto de la propaganda (discontinuidad) obtenida a través del comando `rd`.

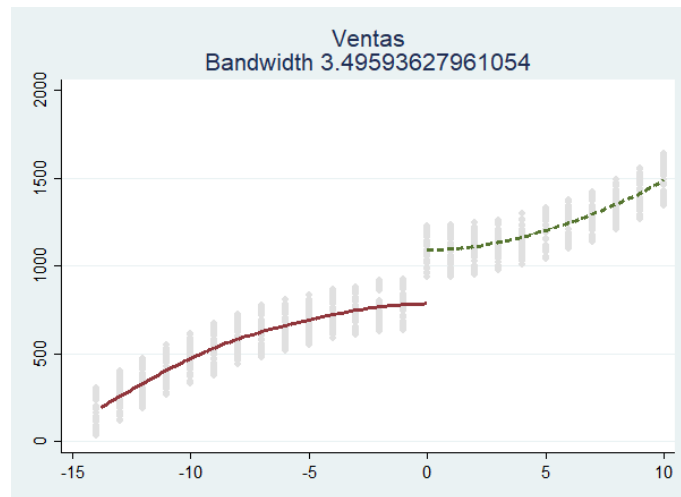


Figura 5: superposición del modelo de regresión discontinua con gráfico de dispersión obtenido con el comando `rd`.

²Código: `rd ventas dif_dias, gr mbw(100) z0(0)`

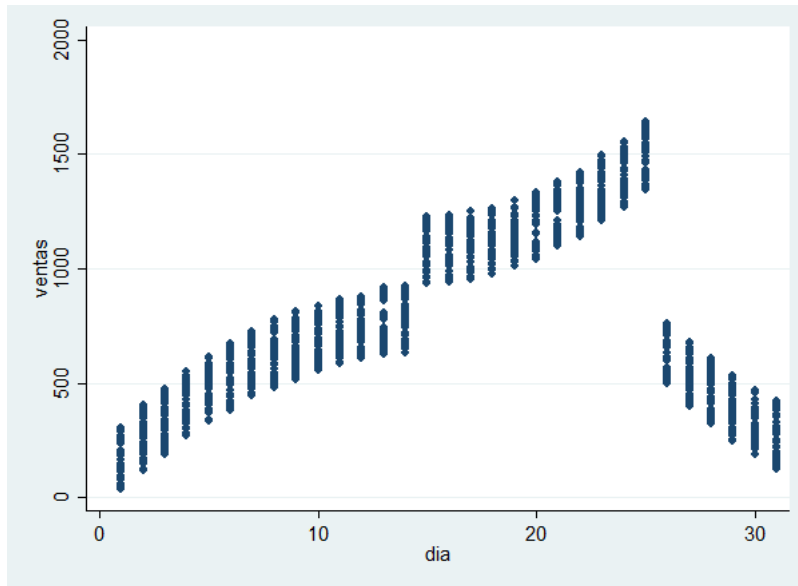
³Código: `rd ventas dia, gr mbw(100) z0(15)`

2. Regresión discontinua (2 discontinuidades)

Esta vez estudiaremos la venta de *spinners* durante todo el mes de diciembre. Para hacer esto, usaremos la base de datos Auxiliar 10 Regresión Discontinua BD 2.dta, la cual llega hasta el día 31. Esta vez supondremos la existencia de 2 discontinuidades: la del día 15 (propaganda, la cual como hemos visto tiene un efecto positivo en las ventas) y la del día 26 (post-navidad, la cual suponemos que tiene un efecto negativo en las ventas).

- a. Obtenga un gráfico de dispersión (o *scatter*) que le permita visualizar cómo se comportó la demanda entre los días 1 y 31 de diciembre.

Respuesta: ver archivo Auxiliar 10 Código.do.



- b. Proponga un modelo de regresión discontinua para dicha demanda. Explique a través de esperanzas condicionales la estructura del modelo.

Respuesta: definiremos los tres tramos que se definen por las dos discontinuidades mediante variables binarias. Sean $T_n^1 = \mathbb{1}_{X_n < 15}$, $T_n^2 = \mathbb{1}_{X_n \in [15, 26)}$ y $T_n^3 = \mathbb{1}_{X_n \geq 26}$. Claramente $T_n^1 + T_n^2 + T_n^3 = 1$.

A continuación se presenta un posible modelo para este comportamiento de demanda y sus respectivas esperanzas condicionales.

$$Y_n = \beta_0 + \gamma_1 T_n^2 + \gamma_2 T_n^3 + \sum_{k=1}^3 \beta_1^k T_n^k X_n + \sum_{k=1}^3 \beta_2^k T_n^k X_n^2 + U_n$$

$$\begin{aligned} \mathbb{E}[Y_n | X_n < 15] &= \beta_0 + \beta_1^1 X_n + \beta_2^1 X_n^2 \\ \mathbb{E}[Y_n | X_n \in [15, 26)] &= \beta_0 + \gamma_1 + \beta_1^2 X_n + \beta_2^2 X_n^2 \\ \mathbb{E}[Y_n | X_n \geq 26] &= \beta_0 + \gamma_2 + \beta_1^3 X_n + \beta_2^3 X_n^2 \end{aligned}$$

A través de las esperanzas condicionales se observa que la estructura de este modelo consiste en 3 funciones cuadráticas distintas para cada tramo. Observando el gráfico de dispersión de la parte anterior, se puede intuir que el modelo –una vez estimado– será creciente y cóncavo en el primer tramo; creciente y convexo en el segundo tramo y decreciente y convexo en el tercer tramo.

- c. Ejecute este modelo de regresión en *Stata* y estime los efectos de la propaganda $\hat{\tau}_1$ (día 15) y del fenómeno post-navidad sobre la demanda $\hat{\tau}_2$ (día 26). ¿Es significativo? Luego sobreponga la predicción al gráfico de dispersión (*scatter*).

Respuesta: ver archivo Auxiliar 10 Código.do. A partir de los datos del cuadro 2, se pueden calcular los efectos inmediatos de propaganda (día 15) y post-navidad (día 26) sobre la demanda (τ)⁴.

VARIABLES	(1) ventas coef	(2) se
propaganda = 2	1,883***	(164.1)
propaganda = 3	4,704***	(1,614)
1b.tramo#c.dia	93.41***	(3.478)
2.tramo#c.dia	-118.0***	(16.66)
3.tramo#c.dia	-232.4**	(113.6)
1b.tramo#c.dia_2	-3.163***	(0.226)
2.tramo#c.dia_2	3.951***	(0.416)
3.tramo#c.dia_2	2.772	(1.992)
Constant	89.31***	(11.34)
Observations	1,550	
R-squared	0.951	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

La primera discontinuidad correspondiente al efecto de la propaganda τ_1 en el día 15 se estima como:

$$\hat{\tau}_1 \approx \underbrace{(89,31 + 1882,9 - 118 \cdot 15 + 3,95 \cdot (15)^2)}_{\lim_{d \rightarrow 15+} \hat{\mathbb{E}}[\text{ventas}|\text{dia}=d]} - \underbrace{(89,31 + 93,4 \cdot 15 - 3,16 \cdot (15)^2)}_{\lim_{d \rightarrow 15-} \hat{\mathbb{E}}[\text{ventas}|\text{dia}=d]} \approx 312,4$$

Como se puede ver en la figura 2, este efecto es estadísticamente significativo.

La segunda discontinuidad correspondiente al efecto post-navidad τ_2 en el día 26 se estima como:

⁴Se usó en lugar de = pues se reemplazaron los valores extraídos del cuadro 2, los cuales están truncados y redondeados por *Stata* y al momento de transcribir el documento. Esto implica leves imprecisiones.

```
. test _cons + 2.tramo + 15 * 2.tramo#dia + 225 * 2.tramo#dia_2 == _cons + 15 * 1.tramo#dia + 225 * 1.tramo#dia_2
> 2

( 1) 2.tramo - 15*1b.tramo#c.dia + 15*2.tramo#c.dia - 225*1b.tramo#c.dia_2 + 225*2.tramo#c.dia_2 = 0

F( 1, 1541) = 454.86
Prob > F = 0.0000
```

Figura 6: test de restricción lineal para evaluar si la discontinuidad dada por $\hat{\tau}_1$ es estadísticamente significativa al 95 % de confianza.

$$\hat{\tau}_2 \approx \underbrace{(89,31 + 4704,38 - 232,42 \cdot 26 + 2,77 \cdot (26)^2)}_{\lim_{d \rightarrow 26+} \hat{\mathbb{E}}[\text{ventas}|\text{dia}=d]} - \underbrace{(89,31 + 1882,9 - 118 \cdot 26 + 3,95 \cdot (26)^2)}_{\lim_{d \rightarrow 26-} \hat{\mathbb{E}}[\text{ventas}|\text{dia}=d]} \approx -950,04$$

Como se puede ver en la figura 2, este efecto es estadísticamente significativo.

```
. test _cons + 3.tramo + 26 * 3.tramo#dia + 676 * 3.tramo#dia_2 == _cons + 2.tramo + 26 * 2.tramo#dia + 676 * 2.
> tramo#dia_2

( 1) - 2.tramo + 3.tramo - 26*2.tramo#c.dia + 26*3.tramo#c.dia - 676*2.tramo#c.dia_2 + 676*3.tramo#c.dia_2 = 0

F( 1, 1541) = 3004.89
Prob > F = 0.0000
```

Figura 7: test de restricción lineal para evaluar si la discontinuidad dada por $\hat{\tau}_2$ es estadísticamente significativa al 95 % de confianza.

Finalmente, se muestra la superposición del modelo de regresión discontinua con gráfico de dispersión.

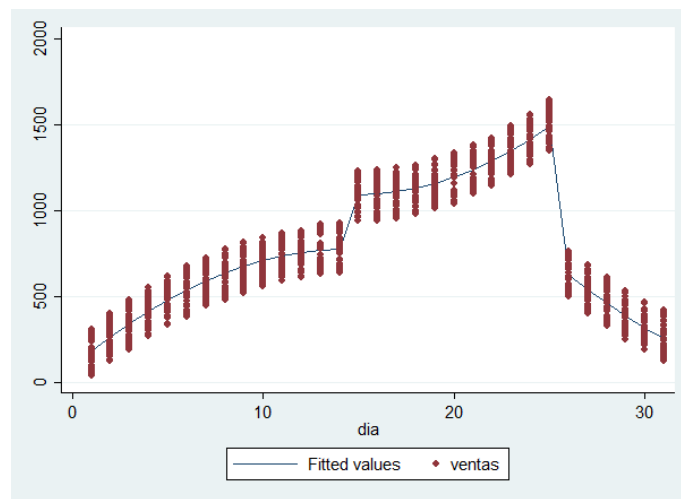


Figura 8: superposición del modelo de regresión discontinua con gráfico de dispersión. La variable independiente es el día.

3. Regresión discontinua difusa

Una vez más estudiaremos la venta de *spinners* durante el mes de diciembre, sin embargo veremos un caso en el que algunas tiendas se adelantaron con el período de propaganda mientras que otras se atrasaron. Dado que el tratamiento (iniciar la propaganda) no coincide estrictamente con la regla de asignación (el hecho de que la propaganda parta supuestamente el día 15), se dificulta el uso de regresión discontinua nítida. Esta vez use la base de datos Auxiliar 10 Regresión Discontinua BD 3.dta.

- a. Obtenga un gráfico de dispersión (o *scatter*) que le permita visualizar cómo se comportó la demanda entre los días 1 y 25 de diciembre.

Respuesta: el gráfico pedido se puede observar en la figura 9. Se observa que en esta nueva base de datos hay tiendas que experimentan el “salto” antes y después de la discontinuidad. Se podría sospechar que estas observaciones corresponden a tiendas que se atrasaron o adelantaron con el comienzo de la propaganda.

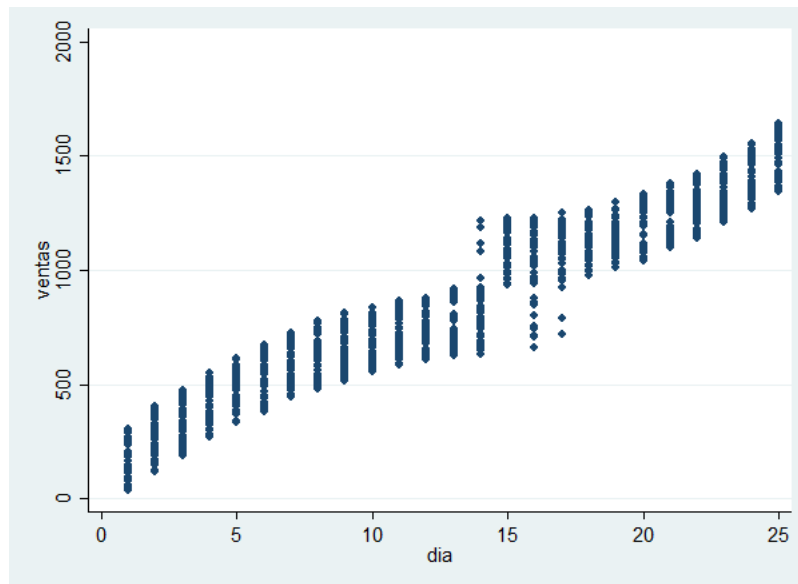


Figura 9: gráfico de dispersión (o *scatter*) para la demanda por *spinners* entre los días 1 y 25 de diciembre

- b. Proponga un modelo de regresión discontinua difusa para dicha demanda, especificando las dos etapas del método.

Respuesta:

- **Primera etapa:** regresionar la variable de tratamiento D_n en función de: (1) un polinomio $g(\cdot)$ de la variable índice X_n desplazada respecto al corte X_c y (2) la variable de asignación. De manera genérica esto se escribe como:

$$D_n = \gamma + g((X_n - X_c))\delta + \lambda A_n + \varepsilon_n$$

En este caso particular D_n es la realización de la propaganda, X_n es el día del mes y A_n es la asignación de la propaganda ($A_n = \mathbb{1}_{X_n \geq 15}$). Por simplicidad consideraremos que $g(\cdot)$ es un

polinomio de grado 1. De esta forma, se propone el modelo para la primera etapa en la cual se usa A_n como instrumento de D_n :

$$D_n = \gamma + \delta_1(X_n - X_c) + \lambda A_n + \varepsilon_n$$

De aquí se obtiene la estimación $\widehat{D}_n$, a partir de la cual se estimará el efecto del tratamiento (propaganda) en la segunda etapa.

- **Segunda etapa:** regresionar la variable dependiente Y_n en función de: (1) un polinomio $g(\cdot)$ de la variable índice X_n desplazada respecto al corte X_c y (2) la estimación $\widehat{D}_n$ obtenida en la etapa 1.

$$Y_n = \alpha + f((X_n - X_c))\beta + \tau\widehat{D}_n + U_n$$

En este caso particular Y_n representa las ventas de *spinners* y el parámetro τ representa el impacto del tratamiento (propaganda), siendo el parámetro de interés. Para lograr un buen ajuste e identificar τ de manera precisa, consideraremos que $f(\cdot)$ es un polinomio de grado 3⁵. De este modo, se propone concretamente el siguiente modelo:

$$Y_n = \alpha + \beta_1(X_n - X_c) + \beta_2(X_n - X_c)^2 + \beta_3(X_n - X_c)^3 + \tau\widehat{D}_n + U_n$$

De aquí se obtiene el estimador $\hat{\tau}$, que representa el impacto de la propaganda sobre las ventas (lo que queríamos medir).

- Ejecute las regresiones correspondientes en *Stata* y estime el efecto de la propaganda τ . ¿Es significativo? Luego sobreponga la predicción al gráfico de dispersión (*scatter*).

Respuesta: ver archivo Auxiliar 10 Código.do. De acuerdo a lo observado en el cuadro que se muestra a continuación, el efecto estimado de las ventas es de 264,5 unidades. El gráfico pedido se encuentra en la figura 10.

⁵No tiene por qué hacerse siempre así. Lógicamente, polinomios de mayor grado lograrán un mejor ajuste. También podría suceder que baste usar un polinomio de grado 1 o 2. Esto dependerá de cómo se comporte la muestra.

	(1)	(2)
	ventas	
VARIABLES	coef	se
dif_dias	20.87***	(1.829)
dif_dias_2	0.521***	(0.0820)
dif_dias_3	0.159***	(0.0115)
Fitted values	264.5***	(16.25)
Constant	800.7***	(10.31)
Observations	1,250	
R-squared	0.941	

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

Cuadro 1: resultados de la estimación mediante regresión discontinua difusa.

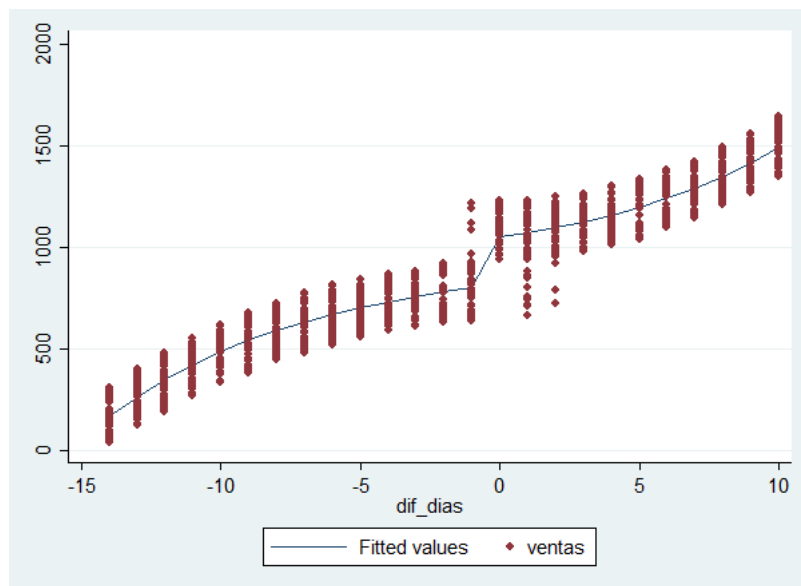


Figura 10: superposición de la predicción con el gráfico de dispersión para las ventas de *spinners* mediante regresión discontinua difusa.