

Econometría Aplicada

Regression Discontinua

Aplicación

Edinson Tolentino

MSc Economics

email: edinson.tolentino@gmail.com

Twitter: [@edutoleraymondi](https://twitter.com/edutoleraymondi)

Educate Peru

4 de diciembre de 2022

- 1 Premier League de Scottish (SPL)
- 2 Pregunta de Investigacion
- 3 Pregunta 1
- 4 Pregunta 2
- 5 Pregunta 3
- 6 Pregunta 4
- 7 Pregunta 5
- 8 Pregunta 6
- 9 Pregunta 7
- 10 Pregunta 8



Contexto para la Premier League de Scottish (SPL)

- La SPL es un ranking actual de 14th en el TOP 55 de las ligas de Europa (Basado sobre la UEFA) y tienen un volumen de ventas que exceden de 200 millones de pound.

- La SPL es un ranking actual de 14th en el TOP 55 de las ligas de Europa (Basado sobre la UEFA) y tienen un volumen de ventas que exceden de 200 millones de pound.
- Acerca de 1/3 de los ingresos del club son originalmente de la asistencia del espectador

- La SPL es un ranking actual de 14th en el TOP 55 de las ligas de Europa (Basado sobre la UEFA) y tienen un volumen de ventas que exceden de 200 millones de pound.
- Acerca de 1/3 de los ingresos del club son originalmente de la asistencia del espectador
- Sobre mucho de los periodos analizados aquí, la SPL ha sido concentrado por dos ciudades grandes de Glasgow (Celtic y Rangers) ambos con una gran información.

- La SPL es un ranking actual de 14th en el TOP 55 de las ligas de Europa (Basado sobre la UEFA) y tienen un volumen de ventas que exceden de 200 millones de pound.
- Acerca de 1/3 de los ingresos del club son originalmente de la asistencia del espectador
- Sobre mucho de los periodos analizados aquí, la SPL ha sido concentrado por dos ciudades grandes de Glasgow (Celtic y Rangers) ambos con una gran información.
- El promedio de asistencia en 2016/2017 temporada fue 14,000

- La SPL es un ranking actual de 14th en el TOP 55 de las ligas de Europa (Basado sobre la UEFA) y tienen un volumen de ventas que exceden de 200 millones de pound.
- Acerca de 1/3 de los ingresos del club son originalmente de la asistencia del espectador
- Sobre mucho de los periodos analizados aquí, la SPL ha sido concentrado por dos ciudades grandes de Glasgow (Celtic y Rangers) ambos con una gran información.
- El promedio de asistencia en 2016/2017 temporada fue 14,000
- Sin embargo, el promedio de asistencia de los juegos de Celtic de Glasgow 3.9 veces el promedio de la SPL, mientras comparable con el caso de Rangers Glasgow 3.5

- Dado la temporada 2001 /2 de la Scottish Premier League (SPL) se comprimió sobre 12 juegos

- Dado la temporada 2001 /2 de la Scottish Premier League (SPL) se comprimió sobre 12 juegos
- Cada equipo inicialmente juega con todos los otros equipos de la League al menos tres veces , obteniendo como resultados 33 juegos

- Dado la temporada 2001 /2 de la Scottish Premier League (SPL) se comprimió sobre 12 juegos
- Cada equipo inicialmente juega con todos los otros equipos de la League al menos tres veces , obteniendo como resultados 33 juegos
- La league entonces se divide a la mitad de seis equipos sobre la base de los resultados luego de los primeros 33 primeros juegos

- Dado la temporada 2001 /2 de la Scottish Premier League (SPL) se comprimio sobre 12 juegos
- Cada equipo inicialmente juega con todos los otros equipos de la League al menos tres veces , obteniendo como resultados 33 juegos
- La league entonces se divide a la mitad de seis equipos sobre la base de los resultados luego de los primeros 33 primeros juegos
- Los top 6 equipos compiten la **Championship League** y la distirbución de seis compiten la **Relegation League**

- Dado la temporada **2001/2** de la SPL contiene 12 equipos

- Dado la temporada **2001/2** de la SPL contiene 12 equipos
- El número de rondas jugadas en la league es 38

- Dado la temporada **2001/2** de la SPL contiene 12 equipos
- El número de rondas jugadas en la league es 38

Number of Rounds	Number of Matches per Round	Number of Seasons (2001/2 to 2017/18)	Number of Match- level Observations
38	6	17	3876
Pre-'split'			
33	6	17	3366
Post-'split'			
5	6	17	510

- Dado la temporada **2001/2** de la SPL contiene 12 equipos
- El número de rondas jugadas en la league es 38

Number of Rounds	Number of Matches per Round	Number of Seasons (2001/2 to 2017/18)	Number of Match-level Observations
38	6	17	3876
Pre-'split'			
33	6	17	3366
Post-'split'			
5	6	17	510

- Luego de **post-split** la información se comprime 510 observaciones usadas para el RDD

- Cuál es el impacto causal sobre la audiencia en el SPL pst-split para un equipo asegure un lugar en la **Chamionship League** comparado con la **Relegation League**

Pregunta 1

- Explique la motivación para usar la Regresión Discontinua (RDD) para esta aplicación.

- Explique la motivación para usar la Regresión Discontinua (RDD) para esta aplicación.
- Cuales son los requisitos estandar para RDD

- Explique la motivación para usar la Regresión Discontinua (RDD) para esta aplicación.
- Cuales son los requisitos estandar para RDD
- Explique su respuesta

Pregunta 1

- Los posición de los clubes son divididos (por ejemplo, despues de 33 juegos). Esto provee un **corte** o **threshold** en este caso.

- Los posición de los clubes son divididos (por ejemplo, despues de 33 juegos). Esto provee un **corte** o **threshold** en este caso.
- El **treatment** es proveido por la exposición para competir en la SPL **Championship League**

- Los posición de los clubes son dividos (por ejemplo, despues de 33 juegos). Esto provee un **corte** o **threshold** en este caso.
- El **treatment** es proveido por la exposición para competir en la SPL **Championship League**
- La division (split) asigna los equipos para un **treatment** y **control** grupos sobre la base una regla deterministica

Pregunta 1

- Los requisitos de la aplicación de una Regression Discontinuity Design (RDD)

- Los requisitos de la aplicación de una Regression Discontinuity Design (RDD)
- **Primero:** la institucion determina la regla para determinar la asignación de tratamiento es pre-determinada y conocida en ventaja al iniciar la temporada.

- Los requisitos de la aplicación de una Regression Discontinuity Design (RDD)
- **Primero:** la institucion determina la regla para determinar la asignación de tratamiento es pre-determinada y conocida en ventaja al iniciar la temporada.
 - ▶ La regla es claramente exogena

- Los requisitos de la aplicación de una Regression Discontinuity Design (RDD)
- **Primero:** la institución determina la regla para determinar la asignación de tratamiento es pre-determinada y conocida en ventaja al iniciar la temporada.
 - ▶ La regla es claramente exógena
- **Segundo:** los administradores de los clubes de la SPL no pueden manipular la regla de división.

- Los requisitos de la aplicación de una Regression Discontinuity Design (RDD)
- **Primero:** la institución determina la regla para determinar la asignación de tratamiento es pre-determinada y conocida en ventaja al iniciar la temporada.
 - ▶ La regla es claramente exógena
- **Segundo:** los administradores de los clubes de la SPL no pueden manipular la regla de división.
- **Tercero:** los clubes de fútbol tienen un impreciso control sobre la posición de la league, ellos logran siempre su posición al final de los 33 juegos, antes de llegar a la división a ser clasificados

- Los requisitos de la aplicación de una Regression Discontinuity Design (RDD)
- **Primero:** la institución determina la regla para determinar la asignación de tratamiento es pre-determinada y conocida en ventaja al iniciar la temporada.
 - ▶ La regla es claramente exógena
- **Segundo:** los administradores de los clubes de la SPL no pueden manipular la regla de división.
- **Tercero:** los clubes de fútbol tienen un impreciso control sobre la posición de la league, ellos logran siempre su posición al final de los 33 juegos, antes de llegar a la división a ser clasificados
- **Cuarto:** es razonable asumir que el producto (asistencia del público) en ausencia del tratamiento (división de los equipos/clubes) podrían mostrar una discontinuidad

Pregunta 2

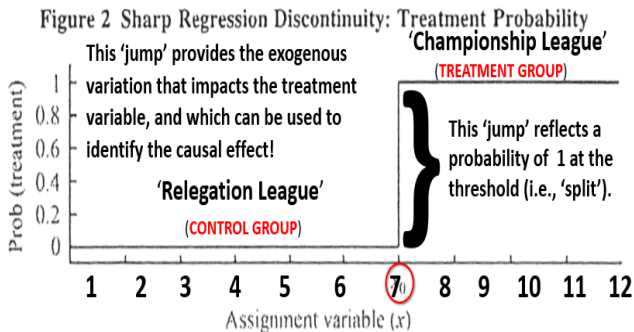
- Qué tipo de regression discontinuity design es?

Pregunta 2

- El RDD en este caso es **sharp** regression discontinuity design

- El RDD en este caso es **sharp** regression discontinuity design
- Existe una regresión discontinua tipo sharp en el tratamiento alrededor del punto de **corte** o valor del **threshold** en la variable del ranking de posición de los clubes.

- El RDD en este caso es **sharp** regression discontinuity design
- Existe una regresión discontinua tipo sharp en el tratamiento alrededor del punto de **corte** o valor del **threshold** en la variable del ranking de posición de los clubes.
- Específicamente existe un salto de probabilidad de los tratados de 0 a 1 alrededor del punto de corte (por ejemplo, ranking de posición de la variable por encima del puesto 7)



A team's league position at the 'split' (i.e., after 33 games)

The rating variable data are recentred to 7 for the empirical analysis.

Pregunta 3

- Describa el atributo clave de la variable ranking de posición. Cuál es la variable producto en este caso?

Pregunta 3

- El ranking (o forcing) es la variable **ordinal** comparado con una variable **continua**

- El ranking (o forcing) es la variable **ordinal** comparado con una variable **continua**
- Esta tiene **6** valores debajo y **6** valores por encima, define el **threshold**

- El ranking (o forcing) es la variable **ordinal** comparado con una variable **continua**
- Esta tiene **6** valores debajo y **6** valores por encima, define el **threshold**
- Una de las ventajas que tiene esta variable discreta ordinal, es que el óptimo del bandwidth (banda) es más precisa de determinar

- El ranking (o forcing) es la variable **ordinal** comparado con una variable **continua**
- Esta tiene **6** valores debajo y **6** valores por encima, define el **threshold**
- Una de las ventajas que tiene esta variable discreta ordinal, es que el óptimo del bandwidth (banda) es más precisa de determinar
- La **variable producto** es el logaritmo de la asistencia del público o espectador.

Pregunta 4

- Construya la variable dummy de tratamiento usando la variable rating y nómbrala como DUM. Tabule el resultado e interprete.

///
gen DUM=0
replace DUM=1 if rating > 6
tab DUM

tab DUM

DUM	Freq.	Percent	Cum.
0	255		
1	255		
Total	510	100.00	

The 'control' group.

The 'treatment' group.

```
///This creates the treatment dummy & tabulates
gen DUM=0
replace DUM=1 if rating > 6
tab DUM
```

- Existen

tab DUM

DUM	Freq.	Percent	Cum.
0	255		
1	255		
Total	510	100.00	

The 'control' group.

The 'treatment' group.

```
///This creates the treatment dummy & tabulates
gen DUM=0
replace DUM=1 if rating > 6
tab DUM
```

tab DUM

DUM	Freq.	Percent	Cum.
0	255		
1	255		
Total	510	100.00	

The 'control' group.

The 'treatment' group.

- Existen

- **255** encuentros de futbol en el **treatment** group (alrededor de izquierda la variable de corte o threshold)

```
///This creates the treatment dummy & tabulates
gen DUM=0
replace DUM=1 if rating > 6
tab DUM
```

tab DUM

DUM	Freq.	Percent	Cum.
0	255		
1	255		
Total	510	100.00	

The 'control' group.

The 'treatment' group.

- Existen

- ▶ **255** encuentros de futbol en el **treatment** group (alrededor de izquierda la variable de corte o threshold)
- ▶ Por otro lado, **255** encuentros de futbol en el **control** group (al lado izquierdo de la variable de corte o threshold)

Pregunta 5

- Usando un grafico de dispersión (scatter plot) examine la relación entre log attendance y rating variable?

- Usando un grafico de dispersión (scatter plot) examine la relación entre log attendance y rating variable?
- Ahora use el comando **rdplot** de STATA, para evaluar si existe evidencia gráfica de una discontinuidad alrededor del threshold

Figure 1: Plot of SPL Log Attendance against League Position

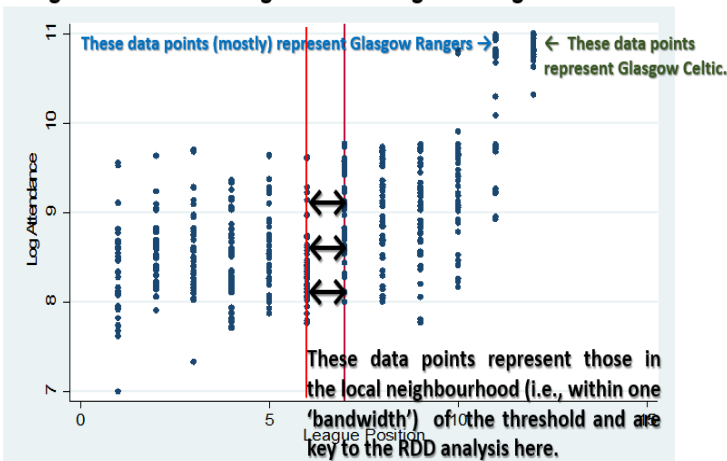


Figure 2: The Regression Discontinuity Plot of SPL Log Attendance at the 'split'

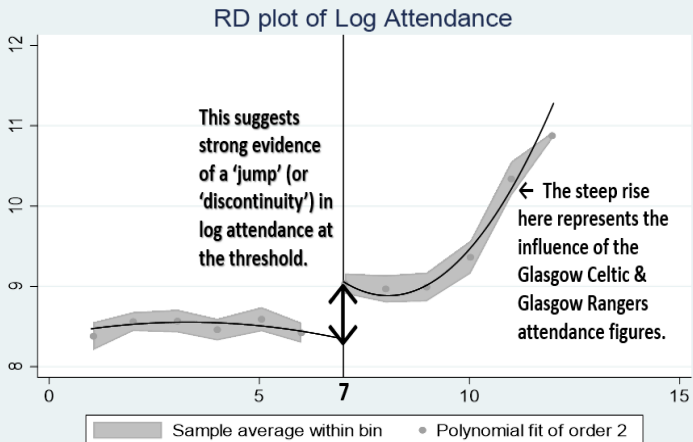
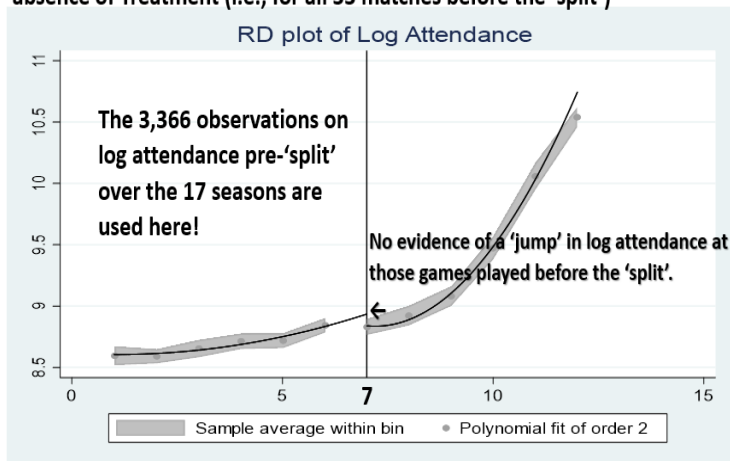


Figure 3: Smoothness across the Threshold in SPL Log Attendance in the absence of Treatment (i.e., for all 33 matches before the 'split')



Pregunta 6

- Usando una **metodología paramétrica** estime la siguiente ecuación por OLS:

$$lattend_i = \beta_0 + \beta_1 (rating - 7) + \beta_2 (rating - 7)^2 +$$

$$+ \beta_3 (rating - 7)_i \times (DUM_i) + \beta_4 (rating - 7)_i^2 \times (DUM_i) + \gamma DUM_i + \epsilon_i$$

- Donde : $i = 1, \dots, 510$

- Usando una **metodología paramétrica** estime la siguiente ecuación por OLS:

$$lattend_i = \beta_0 + \beta_1 (rating - 7) + \beta_2 (rating - 7)^2 +$$

$$+ \beta_3 (rating - 7)_i \times (DUM_i) + \beta_4 (rating - 7)_i^2 \times (DUM_i) + \gamma DUM_i + \epsilon_i$$

- Donde : $i = 1, \dots, 510$
- Explique la propuesta de esta ecuación

- Usando una **metodología paramétrica** estime la siguiente ecuación por OLS:

$$lattend_i = \beta_0 + \beta_1 (rating - 7) + \beta_2 (rating - 7)^2 +$$

$$+ \beta_3 (rating - 7)_i \times (DUM_i) + \beta_4 (rating - 7)_i^2 \times (DUM_i) + \gamma DUM_i + \epsilon_i$$

- Donde : $i = 1, \dots, 510$
- Explique la propuesta de esta ecuación
- Use la significancia al nivel de 0.05 para testear $H_0 : \gamma = 0$ versus $H_a : \gamma \neq 0$

Tabla Estimaciones de lre
regression discontinuity sobre el
logaritmo de asistencia

	Full Sample
Constant	8.3074*** (0.1170)
(rating – 7)	–0.1507* (0.0805)
(rating – 7) ²	–0.0224** (0.0118)
(rating – 7) × DUM	–0.0681 (0.1038)
(rating – 7) ² × DUM	0.1443*** (0.0167)
DUM	0.7176*** (0.1318)
Sample Size	510
Adjusted R ²	0.692

Tabla Estimaciones de lre
regression discontinuity sobre el
logaritmo de asistencia

	Full Sample
Constant	8.3074*** (0.1170)
(rating – 7)	–0.1507* (0.0805)
(rating – 7) ²	–0.0224** (0.0118)
(rating – 7) × DUM	–0.0681 (0.1038)
(rating – 7) ² × DUM	0.1443*** (0.0167)
DUM	0.7176*** (0.1318)
Sample Size	510
Adjusted R ²	0.692

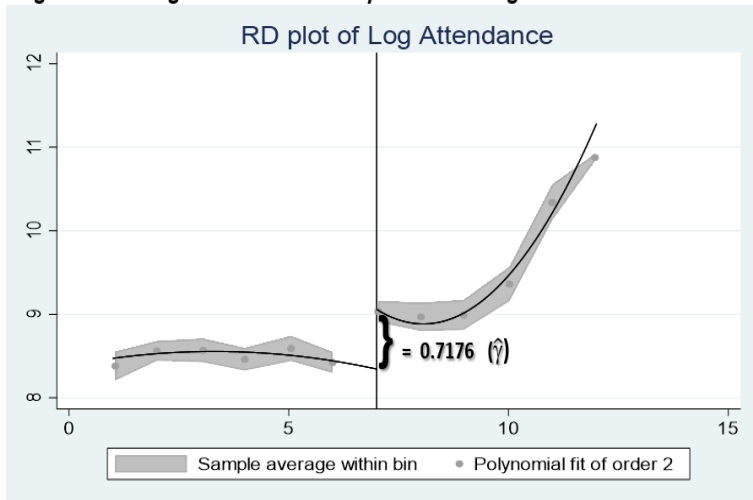
- El efecot estimado para el tratamiento de la
regresión discontinua usando toda la muestra
es de 0.72

Tabla Estimaciones de lre
regression discontinuity sobre el
logaritmo de asistencia

	Full Sample
Constant	8.3074*** (0.1170)
(rating - 7)	-0.1507* (0.0805)
(rating - 7) ²	-0.0224** (0.0118)
(rating - 7) × DUM	-0.0681 (0.1038)
(rating - 7) ² × DUM	0.1443*** (0.0167)
DUM	0.7176*** (0.1318)
Sample Size	510
Adjusted R ²	0.692

- El efecot estimado para el tratamiento de la regresión discontinua usando toda la muestra es de 0.72
- El supuesto indica que, en el threshold, los equipos que califican para **Championship League** section de **SPL** atraen, en promedio , un poc mas del doble de espectadores sobre aquellos equipos ubicados en **Relegation League**.

Figure 2: The Regression Discontinuity Plot of SPL Log Attendance



Pregunta 7

- Estime una regresión no paramétrica lineal con bandwidth de 1.
- Use un nivel de significancia de 0.05 para testear si el efecto del tratamiento es estadísticamente diferente de cero.

Tabla Estimaciones de Ire
regression discontinuity sobre el
logaritmo de asistencia

	Full Sample	Discontinuity Sample
Constant	8.3074*** (0.1170)	8.4227*** (0.8991)
(rating - 7)	-0.1507* (0.0805)	+
(rating - 7) ²	-0.0224** (0.0118)	+
(rating - 7) × DUM	-0.0681 (0.1038)	+
(rating - 7) ² × DUM	0.1443*** (0.0167)	+
DUM	0.7176*** (0.1318)	0.6112*** (0.0891)
Sample Size	510	96
Adjusted R ²	0.692	0.334

Tabla Estimaciones de Ire regression discontinuity sobre el logaritmo de asistencia

	Full Sample	Discontinuity Sample
Constant	8.3074*** (0.1170)	8.4227*** (0.8991)
(rating - 7)	-0.1507* (0.0805)	+
(rating - 7) ²	-0.0224** (0.0118)	+
(rating - 7) × DUM	-0.0681 (0.1038)	+
(rating - 7) ² × DUM	0.1443*** (0.0167)	+
DUM	0.7176*** (0.1318)	0.6112*** (0.0891)
Sample Size	510	96
Adjusted R ²	0.692	0.334

- La muestra de la **discontinuity** es condensada en 96 puntos (observaciones)

Tabla Estimaciones de Ire regression discontinuity sobre el logaritmo de asistencia

	Full Sample	Discontinuity Sample
Constant	8.3074*** (0.1170)	8.4227*** (0.8991)
(rating - 7)	-0.1507* (0.0805)	+
(rating - 7) ²	-0.0224** (0.0118)	+
(rating - 7) × DUM	-0.0681 (0.1038)	+
(rating - 7) ² × DUM	0.1443*** (0.0167)	+
DUM	0.7176*** (0.1318)	0.6112*** (0.0891)
Sample Size	510	96
Adjusted R ²	0.692	0.334

- La muestra de la **discontinuity** es condensada en 96 puntos (observaciones)
- La muestra comprende equipos alrededor de los puntos de la posición del ranking en la 6th y 7th posición

Pregunta 8

- Estime la ecuación agregando las covariables: midweek, ldistance, derby y season
- Use un nivel de significancia de 0.05 para testear si el efecto del tratamiento es estadísticamente diferente de cero.

Tabla Estimaciones de Ire regression discontinuity sobre el logaritmo de asistencia con covariables

	Full Sample	Discontinuity Sample
Constant	57.2213*** (8.7566)	56.9202*** (18.1645)
(rating - 7)	-0.1621** (0.0757)	↓
(rating - 7) ²	-0.0235** (0.0983)	↓
(rating - 7) × DUM	-0.0703 (0.0983)	↓
(rating - 7) ³ × DUM	0.1453*** (0.0161)	↓
DUM	0.7431*** (0.1224)	0.6417*** (0.0857)
midweek	-0.0306 (0.0609)	-0.1275 (0.1109)
ldistance	-0.0482** (0.0234)	-0.0449 (0.0417)
derby	0.3678*** (0.1404)	1.0198*** (0.2264)
season	-0.0242** (0.0044)	-0.0241** (0.0090)
Sample Size	510	96
Adjusted R ²	0.728	0.478

Notes: Robust standard errors in parentheses given the presence of heteroscedasticity; the local linear regression was effectively fitted with a bandwidth of 1; the 'derby' variable captures matches between Glasgow Rangers and Glasgow Celtic, Heart of Midlothian and Hibernian, and Dundee United and Dundee; ↓ not estimated.

Tabla Estimaciones de Ire regression discontinuity sobre el logaritmo de asistencia con covariables

	Full Sample	Discontinuity Sample
Constant	57.2213*** (8.7566)	56.9202*** (18.1645)
(rating - 7)	-0.1621** (0.0757)	↓
(rating - 7) ²	-0.0235** (0.0983)	↓
(rating - 7) × DUM	-0.0703 (0.0983)	↓
(rating - 7) ³ × DUM	0.1453*** (0.0161)	↓
DUM	0.7431*** (0.1224)	0.6417*** (0.0857)
midweek	-0.0306 (0.0609)	-0.1275 (0.1109)
ldistance	-0.0482** (0.0234)	-0.0449 (0.0417)
derby	0.3678*** (0.1404)	1.0198*** (0.2264)
season	-0.0242** (0.0044)	-0.0241** (0.0090)
Sample Size	510	96
Adjusted R ²	0.728	0.478

Notes: Robust standard errors in parentheses given the presence of heteroscedasticity; the local linear regression was effectively fitted with a bandwidth of 1; the 'derby' variable captures matches between Glasgow Rangers and Glasgow Celtic, Heart of Midlothian and Hibernian, and Dundee United and Dundee; ↓ not estimated.

- Los efectos estimados sobre **treatment** alrededor del threshold, usando ambos, completa y la muestra de **discontinuidad**, aparecen invariante para la inclusión de estos cuatro covariables

Tabla Estimaciones de Ire regression discontinuity sobre el logaritmo de asistencia con covariables

	Full Sample	Discontinuity Sample
Constant	57.2213*** (8.7566)	56.9202*** (18.1645)
(rating - 7)	-0.1621** (0.0757)	↓
(rating - 7) ²	-0.0235** (0.0983)	↓
(rating - 7) × DUM	-0.0703 (0.0983)	↓
(rating - 7) ³ × DUM	0.1453*** (0.0161)	↓
DUM	0.7431*** (0.1224)	0.6417*** (0.0857)
midweek	-0.0306 (0.0609)	-0.1275 (0.1109)
ldistance	-0.0482** (0.0234)	-0.0449 (0.0417)
derby	0.3678*** (0.1404)	1.0198*** (0.2264)
season	-0.0242** (0.0044)	-0.0241** (0.0090)
Sample Size	510	96
Adjusted R ²	0.728	0.478

Notes: Robust standard errors in parentheses given the presence of heteroscedasticity; the local linear regression was effectively fitted with a bandwidth of 1; the 'derby' variable captures matches between Glasgow Rangers and Glasgow Celtic, Heart of Midlothian and Hibernian, and Dundee United and Dundee; ↓ not estimated.

- Los efectos estimados sobre **treatment** alrededor del threshold, usando ambos, completa y la muestra de **discontinuidad**, aparecen invariante para la inclusión de estos cuatro covariables
- Esto implica que el **tratamiento** es ortogonal a estas covariables incluidas

Tabla Estimaciones de Ire regression discontinuity sobre el logaritmo de asistencia con covariables

	Full Sample	Discontinuity Sample
Constant	57.2213*** (8.7566)	56.9202*** (18.1645)
(rating - 7)	-0.1621** (0.0757)	↓
(rating - 7) ²	-0.0235** (0.0983)	↓
(rating - 7) × DUM	-0.0703 (0.0983)	↓
(rating - 7) ³ × DUM	0.1453*** (0.0161)	↓
DUM	0.7431*** (0.1224)	0.6417*** (0.0857)
midweek	-0.0306 (0.0609)	-0.1275 (0.1109)
ldistance	-0.0482** (0.0234)	-0.0449 (0.0417)
derby	0.3678*** (0.1404)	1.0198*** (0.2264)
season	-0.0242** (0.0044)	-0.0241** (0.0090)
Sample Size	510	96
Adjusted R ²	0.728	0.478

Notes: Robust standard errors in parentheses given the presence of heteroscedasticity; the local linear regression was effectively fitted with a bandwidth of 1; the 'derby' variable captures matches between Glasgow Rangers and Glasgow Celtic, Heart of Midlothian and Hibernian, and Dundee United and Dundee; ↓ not estimated.

- Los efectos estimados sobre **treatment** alrededor del threshold, usando ambos, completa y la muestra de **discontinuidad**, aparecen invariante para la inclusión de estos cuatro covariables
- Esto implica que el **tratamiento** es ortogonal a estas covariables incluidas
- Existe una tendencia a la baja en la asistencia a la SPL (en las últimas cinco rondas de la liga), lo que puede reflejar una reducción del equilibrio competitivo a largo plazo en la liga.