

Tarea 2

Métodos cuasi-experimentales

Integrantes: Carla Olivares
Javiera Donoso
Daniel Jara
Antonia Arias
Profesor: Daniel Schwartz
Omar David Pérez
Auxiliar:
Francisca Contreras
Karla Toledo Campos
Nicole Troncoso
Rienzi Roldan
Santiago Del Río
Sofía González
Vicente Sule
Fecha de entrega: 7 de Julio de 2023

Métodos Cuasi-Experimentales

Hipótesis

La jubilación y su impacto en el consumo de los hogares ha sido estudiada a nivel mundial. En Italia, se ha investigado este fenómeno conocido como el "Puzzle de consumo tras la jubilación". Según Margherita Borella et al. [1], se observa una caída promedio del 4% en el consumo de bienes no durables como alimentos, productos de limpieza y combustible. Al analizar esta disminución considerando variables como nivel educacional, riqueza y si el retiro es inesperado, los resultados son aún más interesantes. Los sujetos con bajo nivel educacional y baja riqueza muestran una reducción del 8%, pero la caída más significativa ocurre en el grupo de individuos con alto nivel educacional y baja riqueza cuando su retiro es inesperado, alcanzando un 10%.

Otros autores, como Erich Battistin et al.[2], se alejan del análisis descriptivo y exploran las razones detrás de la disminución del consumo de bienes no durables. Concluyen que esta caída no está directamente relacionada con una reducción en los ingresos, sino más bien con una disminución de la necesidad de bienes necesarios para vivir debido al retiro. Además, destacan que la jubilación coincide con el momento en que los hijos abandonan el hogar, lo que también se relaciona con esta reducción. Los resultados al controlar por la cantidad de integrantes confirman este planteamiento, concluyendo que las variables como nivel educacional o riqueza ya no son significativas.

Basándonos en lo mencionado anteriormente, se plantea que la reducción del consumo en los hogares italianos después de la jubilación se debe a la disminución natural de la demanda de bienes no durables debido al retiro, lo cual a su vez, está relacionado con la cantidad de integrantes presentes en el hogar en ese momento.

Preprocesamiento de datos

En una primera parte, se obtuvo los promedios de cada variable, las que se solicitan en el enunciado y de cuatro variables más que se estimen como convenientes. Siguiendo la línea de la hipótesis planteada anteriormente, se decide agregar la variable *ncomp*, la cual es el tamaño de la familia, *educ*, es el nivel educacional recibido por los hombres, *a_r*, que es riqueza real y por último *y_{real}*, que consiste en el ingreso real neto, fijando una base de precios en el año 1995. Además de estas variables, se crea *elegibilidad* y *porcentaje_retirados*, las cuales se obtienen mediante funciones de Rstudio. Una vez que están todas las variables dentro de la base de datos, se obtienen 520 observaciones y 12 variables. Además, se decide sacar los NA para tener datos más puros, y por ende, se tiene un total de 506 observaciones. Por último, cabe destacar que la unidad de análisis que se trabaja con esta base de datos son

personas, específicamente hombres.

Análisis exploratorio de datos

En la figura del anexo 4 se puede observar que en promedio, elegibilidad 0 muestra mayores niveles de consumo que elegibilidad 1, siendo el cambio más significativo con el consumo total, con una diferencia de 0.1. El grupo con elegibilidad 1 muestra en promedio, menores ingresos, mayores niveles de riqueza y menor nivel de educación.

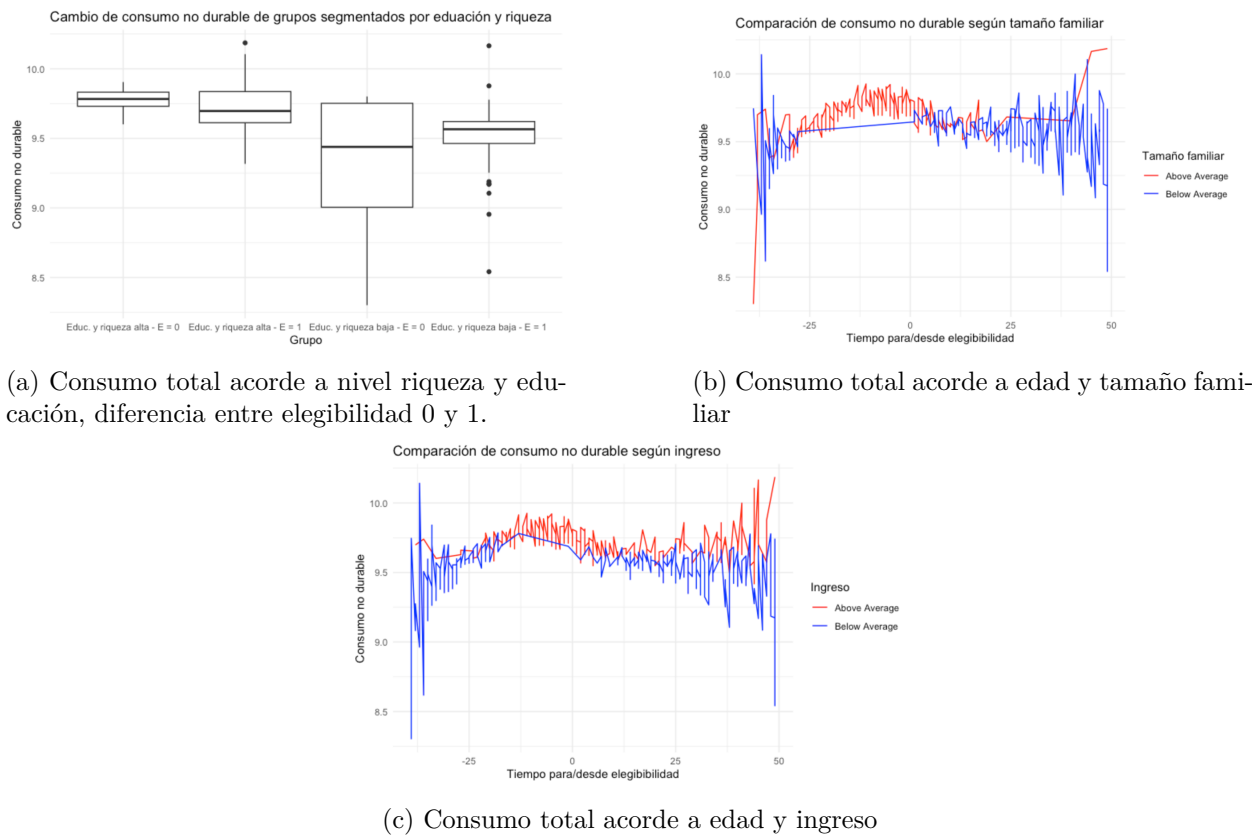


Figura 1: Resultados EDA

Para evaluar lo propuesto en la hipótesis, se crearon dos segmentos en la población, uno con la riqueza y la educación por sobre el promedio, y otro con estos indicadores por debajo del promedio. Para cada grupo se comparó el cambio del consumo total entre aquellos con elegibilidad 0 y 1 (1.a). Resultó que para el grupo con altos niveles de riqueza y educación, el promedio del logaritmo del consumo bajó desde 6.3 a 6.0 aproximadamente, en cambio, la población con bajos niveles presentó un aumento aproximado de 0.1.

Este comportamiento puede atribuirse a la dependencia del consumo de este grupo a otra variable, como por ejemplo el tamaño familiar (con correlación 0.51). También se puede ex-

plicar por la variabilidad de los ingresos de este grupo, la desviación estándar del ingreso es 5.73, pero si se acota para la población mencionada anteriormente, la desviación estándar sube a 6.27.

La segunda componente de la hipótesis es el tamaño familiar, en los gráficos 1.b y 1.c se refleja que tanto el ingreso y el tamaño familiar están por sobre el promedio en los 10 años previos a la edad de jubilación. Acá el consumo total también alcanza sus valores máximos. Luego del periodo de jubilación, el tamaño familiar pasa a ser bajo el promedio, y el consumo pasa a tener una alta variabilidad. Similarmente, ambos segmentos de ingresos tienen consumos variables en este periodo, pero se mantiene la relación de que a mayor ingreso, mayor consumo.

En resumen, de manera general, el consumo total tiene una baja para la edad de jubilación y sí tiene un efecto en el consumo total (correlación 0.96), además el tamaño familiar sí juega un rol.

En las siguientes partes de la tarea, se debe filtrar nuevamente la base de datos utilizando un rango específico para la variable $esse_m$, el cual resulta ser entre $-10 < esse_m < 10$, por lo que se comienza a trabajar con un total de 108 observaciones con 12 variables.

Diseño de Regresión Discontinua

Supuestos

El primer supuesto que debe cumplir el modelo es la discontinuidad del tratamiento en torno a la variable *running*, esto significa que hay un cambio claro en la asignación del tratamiento alrededor del punto de corte de la variable en ejecución. Esto puede ser comprobable mediante un gráfico que observa el comportamiento de la variable $esse_m$ en conjunto con la variable de tratamiento *elegibilidad* y se puede observar como es discontinua cerca del punto de corte. Luego de realizar el gráfico (anexo 6.a) se observa que la variable $esse_m$ sí presenta una discontinuidad en torno a su punto de corte, por lo tanto, el supuesto de discontinuidad se cumple.

El segundo supuesto que debe cumplir el modelo es la no manipulación de la variable *running*, es decir, que los individuos no pueden manipular o controlar su posición dentro de la variable. Esto se puede comprobar con un gráfico que muestre la continuidad o discontinuidad de la curva de densidad de la variable *running*. Al realizar el gráfico (anexo 6.b), se comprueba el supuesto.

El tercer supuesto que el modelo debe cumplir es la continuidad de los resultados potenciales alrededor del punto de corte, es decir que estos resultados relacionados con la variable tratamiento deben permanecer continuos y no mostrar algún cambio brusco en el punto de corte. Este supuesto no es directamente comprobable en la práctica ya que alude a la comparación

de contrafactuales. Es decir, si un individuo que está sobre el puntaje de corte hubiese sido asignado al otro grupo (no hubiese recibido tratamiento) su variable dependiente debería ser muy similar a los que están en el borde del grupo control. Hay que recordar que un supuesto muy importante es que los individuos que están en los bordes de estos grupos son muy similares y su posicionamiento se debe a sucesos aleatorios.

Por último, el cuarto supuesto que debe cumplir el modelo es la continuidad de variables observables alrededor de la punto de corte, es decir, que las características observables y las variables relacionadas con el resultado permanecen continuas y no experimentan cambios repentinos o discontinuidades en el punto de corte. Este supuesto es comprobable realizando regresiones con distinto orden polinomial (anexo 7), a partir de las cuales, analizando la significancia de los coeficientes asociados a la variable $es\text{se}_m$ se puede concluir que no se cumple la continuidad de las variables observables alrededor del puntaje de corte.

Resultados Modelos

Se adjunta la tabla que contiene los 5 modelos solicitados:

Tabla de Regresiones						
Dependent variable:						
	promedio_inc					
	(1)	(2)	(3)	(4)	(5)	(6)
es\text{se}_m	-0.012*** (0.003)	-0.006*** (0.002)	-0.004* (0.002)	-0.005* (0.003)	-0.002 (0.004)	-0.008 (0.013)
elegibilidad	-0.016 (0.029)	-0.007 (0.020)	-0.008 (0.019)	0.005 (0.022)	-0.004 (0.026)	0.015 (0.036)
promedio_yr		0.484*** (0.061)	0.461*** (0.061)	0.501*** (0.070)	0.420*** (0.085)	0.520*** (0.119)
promedio_ar		0.038*** (0.014)	0.051*** (0.015)	0.046*** (0.017)	0.088*** (0.023)	0.088*** (0.029)
promedio_ncomp			0.050** (0.022)	0.048* (0.024)	0.101*** (0.030)	0.097** (0.043)
Constant	9.802*** (0.017)	4.497*** (0.556)	4.432*** (0.545)	4.075*** (0.639)	4.274*** (0.784)	3.257*** (1.110)
Observations	100	100	100	81	58	28
R ²	0.495	0.784	0.795	0.751	0.692	0.769
Adjusted R ²	0.484	0.775	0.784	0.735	0.663	0.716
Residual Std. Error	0.074 (df= 97)	0.049 (df= 95)	0.048 (df= 94)	0.049 (df= 75)	0.048 (df= 52)	0.046 (df= 22)
F Statistic	47.473*** (df= 2; 97)	86.443*** (df= 4; 95)	73.038*** (df= 5; 94)	45.347*** (df= 5; 75)	23.400*** (df= 5; 52)	14.644*** (df= 5; 22)

Note:

*p<0.1; **p<0.05; ***p<0.01

Figura 2: Modelos RD

Conclusiones

Al analizar la tabla anterior, se puede observar un factor muy importante, el cual es que la variable *elegibilidad* no es significativa, lo cual podría venir dado por la relación que existe entre *es\text{se}_m* y *elegibilidad*. Sin embargo, se puede concluir que el consumo de las personas al momento de jubilarse disminuye, dado el signo de la variable *elegibilidad* y también por el alto poder estadístico que tiene esta misma, siguiendo la tendencia que se plantea en el

análisis exploratorio de datos, en donde se observa este mismo comportamiento. Además, puede observarse que no existe un efecto sobre el promedio del ingreso real neto y sobre el promedio de la riqueza acumulada, al momento de controlar por el promedio del tamaño familiar, es decir que a diferencia de lo que plantea Battistin (2008) en sus estudios, la riqueza acumulada y los ingresos siguen siendo variables relevantes, sin importar si se considera o no la cantidad de hijos, que en este caso, se expresa mediante la disminución del tamaño familiar. Por lo tanto, no hay evidencia suficiente para comprobar la hipótesis planteada en un inicio, por lo cual según estos modelos se rechaza. Cabe mencionar, que se puede suponer que el rechazo de esta hipótesis se debe principalmente a la falta de datos de la base en la cual se está trabajando.

Diseño de Variables Instrumentales

Supuestos

El primer supuesto que debe cumplir el modelo es la relevancia del instrumento, la cual es verificable a través del test de hipótesis y además se puede comprobar mediante la covarianza entre la variable Z , la cual en este caso sería *elegibilidad* en conjunto con la variable X , donde en este estudio es *porcentaje_retirados*, es decir, $Cov(Z, X) \neq 0$. Se puede observar que la covarianza es 0.14 (anexo 8.b), por lo tanto, el supuesto se cumple.

Luego, el segundo supuesto que debe cumplir el modelo es la independencia de la variable instrumental con respecto al error del modelo. Este supuesto se puede comprobar mediante una regresión lineal con *elegibilidad* y las covariables del modelo. Como se puede observar en la tabla 9 del anexo, para *promedio_yr* y *promedio_ncomp* el supuesto no se cumple, ya que sus b etas son significativos. El supuesto solo se cumple para la covariable *promedio_ar*. Como no se cumple en general para todas las covariables, se concluye que el supuesto 2, no se cumple. Por último, el tercer supuesto que debe cumplir el modelo es la restricción de exclusión, donde se puede comprobar mediante las explicaciones de las variables Z , X e Y :

Z = Es una variable binaria, toma el valor 1 si el jefe del hogar ya tiene la edad mínima para jubilarse, y 0 en otro caso.

X = Es el porcentaje de retirados promedio del estatus de jubilación de los hombres (trabajadores, retirado u otro).

Y = Es el gasto total de las personas que participaron del modelo.

El supuesto nos dice que tiene que existir una única relación entre la variable Y con X , y que no puede existir una relación directa con la variable instrumental Z . Como Y es el gasto total de las personas, no puede estar relacionado de manera directa con si el jefe de hogar tiene la edad mínima para jubilarse o no, ya que de la única forma que pueden estar relacionados, es que a partir del gasto total de las personas, se pueda obtener el porcentaje

de retirados promedio y con ello, poder relacionarlo con la variable instrumental Z y como esto no es posible, el supuesto sí se cumple.

Resultados Modelos

Tabla de Regresiones						
Dependent variable:						
	(1)	(2)	(3)	(4)	(5)	(6)
porcentaje_retirados	-0.213*** (0.022)	-0.100*** (0.033)				
promedio_ar		0.059*** (0.014)		0.056*** (0.015)		0.058*** (0.015)
promedio_ncomp		0.028 (0.026)		0.070*** (0.019)		0.036 (0.032)
promedio_yr		0.434*** (0.062)		0.473*** (0.061)		0.444*** (0.067)
eligibilidad			-0.128*** (0.016)	-0.029** (0.015)		
X_hat					-0.216*** (0.027)	-0.088** (0.044)
Constant	9.883*** (0.012)	4.722*** (0.572)	9.861*** (0.011)	4.206*** (0.534)	9.884*** (0.014)	4.602*** (0.647)
Observations	100	100	100	100	100	100
R ²	0.479	0.799	0.393	0.789	0.393	0.789
Adjusted R ²	0.474	0.791	0.387	0.780	0.387	0.780
Residual Std. Error	0.074 (df = 98)	0.047 (df = 95)	0.080 (df = 98)	0.048 (df = 95)	0.080 (df = 98)	0.048 (df = 95)
F Statistic	90.092*** (df = 1; 98)	94.560*** (df = 4; 95)	63.580*** (df = 1; 98)	88.897*** (df = 4; 95)	63.580*** (df = 1; 98)	88.897*** (df = 4; 95)
Note:					*p<0.1; **p<0.05; ***p<0.01	

Figura 3: Regresiones variable instrumental

Conclusiones

En la figura 3 se estudia la relación entre la proporción de jubilados y el consumo. Según lo visto en el primer modelo, se deduce que existe una relación entre ambas variables. El coeficiente, que es significativo al 1 %, muestra que para un aumento de un 10 % en el porcentaje de retirados, se estaría reduciendo el consumo en un 2,13 %. Sin embargo, al ver el modelo 2, pareciera que gran parte de ese efecto proviene de otras fuentes, como la riqueza y el ingreso.

Luego, utilizando la variable elegibilidad, se intenta realizar un análisis que aísla a la relación que tiene *porcentaje_retirados* con variables no observadas (modelo 5 y 6). Este análisis se realiza con el método de mínimos cuadrados en dos etapas, es decir, primero se estimó el porcentaje de retirados (X_{hay}) a través de la variable elegibilidad y luego se estima el consumo a partir de el porcentaje de retirados estimado. Los resultados no muestran un cambio significativo cuando no se controla por las covariables, sin embargo, cuando éstas son agregadas el coeficiente asociado a X_{hat} aumenta a -0.88, en contraste con el -0.1 que se tenía en el segundo modelo.

También se estimó la intención de tratar con modelos que estiman el consumo a partir de la variable elegibilidad. Aquí pareciera que el efecto de elegibilidad es alto por el valor -0.128 que se aprecia en el modelo (3), el cual se interpreta como una reducción de un 12.8% del consumo. Sin embargo, nuevamente agregar las variables de control muestra que el efecto realmente es de otras variables. En el caso del modelo (4) el coeficiente asociado a elegibilidad es -0.029 que se interpreta como una caída del 2.9% .

Conclusiones

Luego de analizar los datos proporcionados para esta investigación, con el fin de evaluar si existe suficiente evidencia para aceptar la hipótesis inicial, se pueden evidenciar diferentes aspectos interesantes de mencionar. Primero que todo, la base de datos, luego de los ajustes realizados, queda con una cantidad muy pequeña de datos, lo cual puede perjudicar la robustez del estudio, lo cual se observa al realizar los modelos y posteriores análisis. De los primeros 5 modelos realizados puede concluirse que efectivamente, faltan datos para aprobar la hipótesis por lo cual esta se rechaza, mostrando que la riqueza e ingresos siguen siendo significativos a pesar de que se controle por el tamaño del grupo familiar. También es importante mencionar que no todos los supuestos se cumplen, ni para regresión discontinua ni para variable instrumental, por lo que asumir que los supuestos se cumplen para los posteriores análisis, también le quita validez interna al estudio.

Por otra parte, se considera que al ser los datos acotados a un país en específico y a solo ciertos años, se pierde la validez externa del estudio, debido a que la realidad económica varía bastante de un país a otro, por lo que los resultados podrían ser distintos para otros países, como por ejemplo para un país de Latinoamérica.

Con respecto a la validez del constructo, en el estudio se asume que todas las personas realmente se jubilan a la edad asignada para la jubilación, lo cual en la realidad no es necesariamente cierto, ya que existe un gran porcentaje de personas que siguen trabajando luego de jubilarse, lo cual hace que se pierda la validez del constructo en el análisis.

Finalmente, es necesario mencionar la importancia de realizar análisis y estudios con la cantidad suficiente de datos para llegar a resultados robustos y válidos para así, poder encontrar hallazgos y conclusiones relevantes, aplicables a otras situaciones.

Referencias

[1] Borella, M., Coda Moscarola, F., y Rossi, M., “(un) expected retirement and the consumption puzzle,” Empirical Economics, vol. 47, pp. 733–751, 2014.

[2] Battistin, E., Brugiavini, A., Rettore, E., y Weber, G., “The retirement consumption puzzle: evidence from a regression discontinuity approach,” American Economic Review, vol. 99, no. 5, pp. 2209–2226, 2009.

EDA

	elegibilidad<dbl>	promedio_inc<dbl>	promedio_lncn<dbl>	promedio_lnjconsal<dbl>	promedio_yr<dbl>	promedio_ar<dbl>	promedio_educ<dbl>	promedio_acontrib<dbl>
0		9.751399	9.650388	5.973016	26004.12	141552.6	3.787790	16.62608
1		9.636749	9.595252	5.967641	25346.66	150453.5	3.012228	37.38774

Figura 4: Estadísticos importantes

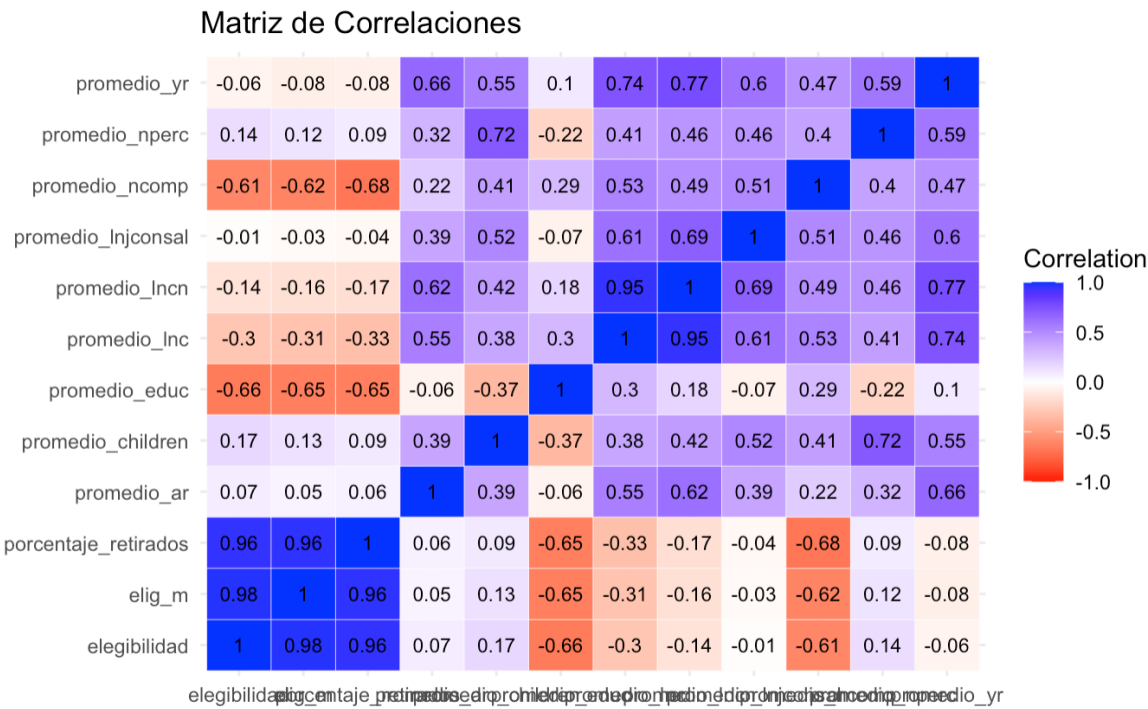
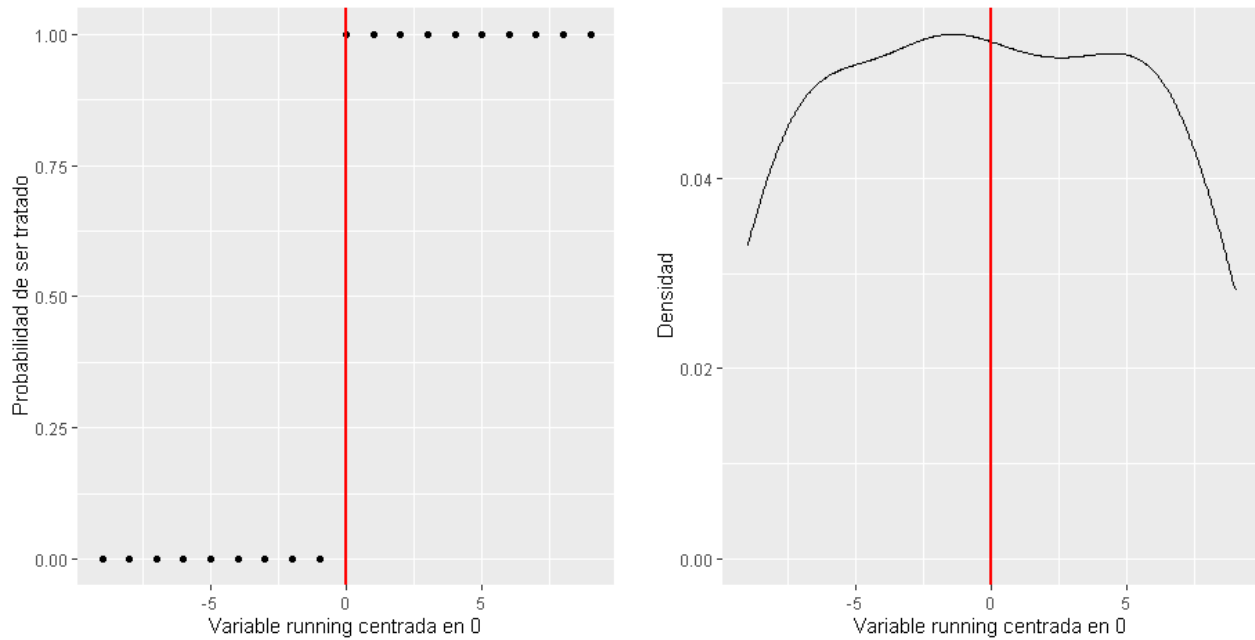


Figura 5: Correlaciones

Regresión Discontinua



(a) Primer Supuesto RD

(b) Segundo Supuesto RD

Figura 6: Supuestos RD

Tabla de Regresiones ncomp									
	Dependent variable:								
	promedio_ncomp			promedio_ar			promedio_yr		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
elegibilidad	-0.053 (0.094)	-0.051 (0.095)	0.023 (0.126)	0.190 (0.165)	0.174 (0.162)	-0.005 (0.215)	-0.034 (0.038)	-0.039 (0.036)	-0.059 (0.048)
esse_m	-0.054*** (0.009)	-0.054*** (0.009)	-0.072*** (0.022)	-0.019 (0.015)	-0.019 (0.015)	0.024 (0.037)	-0.010*** (0.003)	-0.010*** (0.003)	-0.005 (0.008)
I(esse_m2)		0.001 (0.001)	0.001 (0.001)		-0.003** (0.002)	-0.004** (0.002)		-0.001*** (0.0003)	-0.001*** (0.0004)
I(esse_m3)			0.0002 (0.0003)			-0.001 (0.0004)			-0.0001 (0.0001)
Constant	3.221*** (0.055)	3.205*** (0.062)	3.163*** (0.077)	10.789*** (0.096)	10.896*** (0.106)	10.997*** (0.132)	10.118*** (0.022)	10.153*** (0.024)	10.164*** (0.030)
Observations	100	100	100	100	100	100	100	100	100
R ²	0.643	0.644	0.647	0.016	0.063	0.079	0.357	0.421	0.423
Adjusted R ²	0.635	0.633	0.632	-0.004	0.034	0.040	0.343	0.402	0.399
Residual Std. Error	0.237 (df= 97)	0.237 (df= 96)	0.238 (df= 95)	0.416 (df= 97)	0.408 (df= 96)	0.407 (df= 95)	0.095 (df= 97)	0.090 (df= 96)	0.091 (df= 95)
F Statistic	87.219*** (df= 2; 97)	57.837*** (df= 3; 96)	43.495*** (df= 4; 95)	0.805 (df= 2; 97)	2.153* (df= 3; 96)	2.027* (df= 4; 95)	26.882*** (df= 2; 97)	23.220*** (df= 3; 96)	17.413*** (df= 4; 95)

Note:

*p<0.1; **p<0.05; ***p<0.01

Figura 7: Cuarto supuesto: Regresiones

Variables instrumentales

```
call:
lm(formula = x ~ Z, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-0.49258 -0.06686 -0.00412  0.08823  0.29041

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.10243    0.02035   5.035 1.97e-06 ***
Z            0.59327    0.02851  20.809 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1481 on 106 degrees of freedom
Multiple R-squared:  0.8033,    Adjusted R-squared:  0.8015
F-statistic: 433 on 1 and 106 DF, p-value: < 2.2e-16
```

(a) Regresión Lineal Supuesto IV

```
> cov(X,Z)
[1] 0.1496522
```

(b)

Figura 8: Covarianza entre X y Z

Tabla de Regresiones			
	Dependent variable:		
	promedio_yr (1)	promedio_ar (2)	promedio_ncomp (3)
eligibilidad	-0.128*** (0.020)	0.009 (0.083)	-0.553*** (0.055)
Constant	10.167*** (0.014)	10.885*** (0.060)	3.485*** (0.040)
Observations	100	100	100
R ²	0.301	0.0001	0.503
Adjusted R ²	0.294	-0.010	0.498
Residual Std. Error (df = 98)	0.098	0.417	0.277
F Statistic (df = 1; 98)	42.283***	0.012	99.374***
Note: * p<0.1; ** p<0.05; *** p<0.01			

Figura 9: Tabla regresiones de covariables contra elegibilidad