

Práctica Calificada: Regression Discontinuity

Profesor: Tomás Rau

Esta tarea podrá ser respondidas en grupos de hasta 5 personas. Se pide elaborar un informe con las respuestas y adjuntar el código en STATA o R.

La base de datos **almond.dta** es una submuestra de los datos utilizados en el paper *Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns* de Almond, Doyle, Kowalski & Williams (2010); de modo que muchos de los resultados que replique no serán exactos, sino más bien parecidos.

- a) Lea el paper y conteste las siguientes preguntas:
- a) ¿Por qué podría haber una discontinuidad en el estado de salud de los recién nacidos justo debajo vs justo encima de 1500gr?
 - b) ¿Cuál es el supuesto fundamental que se hace sobre la salud de los recién nacidos cerca del corte *Very Low Birth Weight* que permite a los autores estimar el retorno marginal de los cuidados médicos?
 - c) ¿Qué podemos decir sobre el supuesto a medida que nos alejamos (por ambos lados) del peso de corte?
 - d) ¿Cuáles son las 2 formas en las que los autores miden inputs sobre el estado de salud?
 - e) ¿De qué forma miden el output o resultado sobre el estado de salud?
 - f) ¿Es un diseño Sharp o Fuzzy? Justifique.
- b) Análisis preliminar: los autores proponen 2 formas de estimar la discontinuidad del outcome en 1500gr: una *Local Linear Regression* y un modelo lineal controlando por covariables y con efectos fijos. Comencemos con el modelo lineal. Para todas las preguntas, ignore y omita los efectos fijos y los covariables de la ecuación 1 del paper.
- a) Escriba la forma funcional del valor esperado de Y para recién nacidos con un peso de (i) 1450, (ii) 1550, (iii) 1499 y (iv) 1501 gramos. Nota: no necesita estimar nada ni utilizar datos
 - b) ¿Cómo interpreta los coeficientes α_2 y α_3 , y por qué podrían ser distintos?
 - c) En el punto de corte hay 2 interceptos, uno para aquellos justo debajo de la clasificación VLBW y el otro para los recién nacidos justo encima. Escriba la forma funcional de ambos interceptos e interprete la diferencia (i.e. la resta de ambos interceptos).
- c) Análisis preliminar 2: Piense ahora en la *Local Linear Regression*:
- a) ¿Cuál es la intuición de una regresión Local?
 - b) ¿Por qué es importante la elección del ancho de banda? Explique el trade-off sesgo-varianza.
 - c) ¿Cómo elegir entonces un ancho de banda óptimo? Explique brevemente los métodos propuestos por Imbens and Kalyanaraman (2012) y Calonico, Cattaneo, and Titiunik (2015).
 - d) ¿Qué rol juega la elección del Kernel? Explique brevemente la diferencia entre Kernel Uniforme, Triangular y Epanechnikov.
- d) Estime el ATE en el punto de corte utilizando el modelo lineal, interprete sus resultados, compárelos con los obtenidos por los autores y construya una gráfica de la esperanza condicional de Y en el peso al nacer. Nota: Incluya la nube de puntos (scatter) en su gráfico.
- e) Estime el ATE en el punto de corte utilizando un *Local Linear Regression* (comando *rdrobust*). Pruebe con distintas especificaciones:

- Para un ancho de banda dado (CCT), estime con Kernel Uniforme, Triangular y Epanechnikov. Comente sus resultados
 - Para un Kernel dado (Epanechnikov), estime con CCT, IK, $2 \cdot \text{CCT}$ y $0.5 \cdot \text{CCT}$. Comente sus resultados y compárelos con los obtenidos por Almond et al (2010).
- f) Utilice el comando *rdplot* para graficar la relación no lineal entre la esperanza condicional de Y en el peso al nacer y compare con la gráfica construida en d).
- g) Construya el histograma del peso al nacer, realice el test de McCrary y discuta sus resultados. ¿Cuál es el argumento de los autores para justificar que la discontinuidad de la densidad no representa una amenaza a la identificación? Nota: debe instalar manualmente el test en Stata desde la página web del autor.
- h) Lea el paper de Barreca, Guldi, Lindo & Waddell (2011) y explique brevemente el argumento de los autores para revisar los resultados de Almond et al (2010).
- i) Barreca et al (2011) proponen una estimación alternativa conocida como *Donut-hole RD*. Estime el ATE en el punto de corte con una *Local Linear Regression*, (i) eliminando las observaciones con peso de 1500 y (ii) eliminando las observaciones con peso entre 1499 y 1501. Comente sus resultados y compárelos con los obtenidos por Barreca et al (2011).