

## Session 3 - RDD

### Para esta sesión:

Replicaremos algunos de los resultados del siguiente artículo: Ozier, Owen. (2018). "The Impact of Secondary Schooling in Kenya: A Regression Discontinuity Analysis," *Journal of Human Resources*, 53(1), 157-188.

### Preguntas Conceptuales

**Contexto.** La pregunta principal de investigación que el autor, Ozier (2015), pretende responder es: ¿Cuáles son los impactos de la educación secundaria en la acumulación de capital humano, la elección ocupacional y la fertilidad de los jóvenes adultos en Kenia? Esta pregunta de investigación tiene implicaciones políticas significativas, particularmente para las estrategias de desarrollo educativo y económico en Kenia. Si la educación secundaria afecta positivamente el capital humano y las elecciones ocupacionales, las políticas que promuevan un acceso más amplio a la educación secundaria podrían mejorar el desarrollo económico y los resultados de vida individuales. Además, la reducción de las tasas de embarazo adolescente entre las mujeres podría tener implicaciones políticas a largo plazo relacionadas con la demografía y la salud, influyendo en las decisiones relacionadas con la planificación familiar y los programas educativos.

1. El autor utilizó un diseño de discontinuidad de regresión porque creía que una especificación OLS simple sería insuficiente. Considera el efecto de asistir a cualquier escuela secundaria en uno de los resultados de interés (nivel educativo alcanzado, tasa de autoempleo de baja habilidad, fertilidad). ¿Cuáles son dos posibles variables omitidas (confusores) que sesgarían los resultados de una especificación OLS simple? Explica el mecanismo de la variable omitida y usa la fórmula de sesgo de variable omitida para argumentar si llevaría a una subestimación o sobreestimación del efecto verdadero.
2. ¿Por qué es importante probar la continuidad de las características observables antes del tratamiento a través del corte de puntuación de la prueba?
3. Explica el propósito de la Figura 4. ¿Cómo se compara con la Figura 6? Explica cómo se construyen ambas figuras.

### Preguntas de Análisis de Datos

#### Principales variables de interés:

- `kcpe_self_or_matched_recent`: Puntuación más reciente del KCPE, autoinformada, corregida si los datos administrativos están disponibles (escala de 500)
- `finishsecondary`: Indicador de completar al menos 12 años de escolaridad (escuela secundaria)
- `has_score_2016`: Indicador de reportar una puntuación (posible y válida) del KCPE

- `rkcpe`: Puntuación del KCPE (primer intento, datos administrativos confirmados), reescalada y recentrada en los cortes relevantes  $((KCPE - \text{corte})/100)$
  - `passrkcp`: Indicador de si la puntuación del KCPE (primer intento, datos administrativos confirmados, recentrada en los cortes) supera el corte relevante
  - `int_pass_rkcpe`: Interacción entre `rkcpe` y `passrkcp`
  - `female`: Indicador de si el encuestado es mujer
  - `ravens_plus_vocab_standardized`: Suma estandarizada de las puntuaciones estandarizadas en las pruebas de Ravens B y Vocabulario
1. Crea estadísticas resumidas para `kcp_self_or_matched_recent`, `finishsecondary`, y otras dos variables de tu elección de la Tabla 1. Para coincidir con la Tabla 1, restringe las observaciones con una puntuación válida de KCPE (`has_score_2016 == 1`), y para las variables de resultado, además restringe a un ancho de banda de 80 puntos alrededor del corte de puntuación (el valor absoluto de `rkcpe`  $\leq$  0.8). Ten en cuenta que los Paneles D y E requieren restricciones adicionales.
  2. Crea una tabla que ilustre el efecto de primer nivel del corte de puntuación de la prueba (`passrkcp`) sobre la probabilidad de completar la escuela secundaria (`finishsecondary`). Replica los coeficientes y errores estándar de las Columnas 1, 4 y 7 de la Tabla 2, las estimaciones de primer nivel para las tres muestras diferentes (Agrupadas, Masculinas y Femeninas) sin controles. Usa un modelo lineal, un núcleo uniforme y un ancho de banda de 80 puntos alrededor del corte de puntuación (establece el valor absoluto de `rkcpe`  $\leq$  0.8). Agrupa en el nivel de puntuación de la prueba (`rkcpe`), pero ten en cuenta que está en debate cuándo se deben agrupar los errores estándar en los diseños de discontinuidad de regresión.
  3. Crea un gráfico de discontinuidad de regresión usando una aproximación polinomial lineal, para ilustrar el efecto de superar el corte en las puntuaciones cognitivas en la adultez (`ravens_plus_vocab_standardized`). Para replicar el panel B de la Figura 6, usa datos dentro de un ancho de banda de 80 puntos del corte de puntuación, y usa intervalos uniformemente espaciados de 10 puntos.