

Modelación en Ciencia de Datos

Sesión 09

Matching

Referencia: Dehejia, Rajeev H., and Sadek Wahba. 2002. *Propensity Score-Matching Methods for Nonexperimental Causal Studies*. Review of Economics and Statistics 84 (1): 151–61.

En esta práctica analizaremos el artículo de Dehejia y Wahba (2002), y replicamos algunos de sus principales resultados. Los autores utilizan la metodología de *matching* para examinar el efecto de un programa de capacitación laboral en EUA (*NSW job training program*). Para realizar este análisis, utilizaremos el conjunto de datos adjunto a su artículo. Este conjunto de datos está disponible en Canvas y fue tomado de:

<https://mixtape.scunning.com/matching-and-subclassification.html>

1. Revise la descripción del programa que se realiza en la Sección I y replique la Tabla 1 de la muestra utilizada por los autores para describir a su conjunto de datos.
2. Calcule el *ATE* del experimento. Deberá obtener que el efecto del NSW job-training program sobre los ingresos reales fue un incremento de \$1,794.343. (renglón 1 de la tabla 2).
3. Siguiendo el procedimiento de los autores, ahora utilice la información de la encuesta CPS como grupo de control no experimental. Agregue este conjunto de datos a los datos experimentales y estime el *propensity score* usando un modelo logit.
4. Mediante la construcción del histograma para ambos grupos, analice la región del *commun support*
5. Utilice el procedimiento de *matching* bajo los siguientes criterios para calcular el ATE:
 - a) Weighting on the propensity score
 - b) Nearest-neighbor matching
 - c) Coarsened exact matching

Compare sus resultados y concluya el efecto del tratamiento en el contexto del problema.