

Tarea 2 - Evaluación de Impacto

CIA, Regresión y Matching

Fecha de Entrega: 14 de mayo, 2025

Instrucciones

- Debe entregar esta tarea a más tardar el día 14 de mayo a las 11:59 pm.
- La entrega se realiza via Canvas en la sección de Tareas.
- Debe entregar dos archivos:
 1. El .pdf con las respuestas, tablas y figuras y los análisis solicitados, generado con LaTeX, Quarto o RMarkdown.
 2. El código que genera los análisis, ya sea .do, .r, .rmd, o .qmd.
- Mi recomendación es usar R + Quarto markdown (o su versión previa, RMarkdown):
 - Puede revisar el siguiente tutorial para iniciarse en Quarto: <https://quarto.org>.
 - Asegúrese de usar formato pdf como formato de salida.

Descripción de la Tarea

En esta tarea buscamos replicar parte de los resultados de Dehejia & Wahba (1999). Como vimos en clase, este paper se evalúa el desempeño de estimaciones no experimentales (basadas en el *supuesto de independencia condicional (CIA)*) del efecto causal del programa de capacitación laboral NSW. Para esto, el trabajo utiliza estimaciones experimentales como *benchmark*.

Los datos se pueden obtener directamente de los siguientes links:

- Datos de la evaluación experimental: https://raw.githubusercontent.com/scunning1975/mixtape/master/nsw_mixtape.dta
- Datos de la encuesta CPS para la construcción del grupo de comparación no experimental: https://raw.githubusercontent.com/scunning1975/mixtape/master/cps_mixtape.dta

NSW job training program

El paper de Dehejia y Wahba (1999) es una reevaluación de un paper anterior de Lalonde (1986), el que vieron en la ayudantía. Ambos papers evalúan el programa de capacitación laboral NSW. Para mayor contexto, transcribo acá la descripción del programa que hace Scott Cunningham en su libro “Causal Inference: The Mixtape”:

The National Supported Work Demonstration (NSW) job-training program was operated by the Manpower Demonstration Research Corp (MRDC) in the mid-1970s. The NSW was a temporary employment program designed to help disadvantaged workers lacking basic job skills move into the labor market by giving them work experience and counseling in a sheltered environment. It was also unique in that it randomly assigned qualified applicants to training positions. The treatment group received all the benefits of the NSW program. The controls were basically left to fend for themselves. The program admitted women receiving Aid to Families with Dependent Children, recovering addicts, released offenders, and men and women of both sexes who had not completed high school.

Treatment group members were guaranteed a job for nine to eighteen months depending on the target group and site. They were then divided into crews of three to five participants who worked together and met frequently with an NSW counselor to discuss grievances with the program and performance. Finally, they were paid for their work. NSW offered the trainees lower wages than they would've received on a regular job, but allowed for earnings to increase for satisfactory performance and attendance. After participants' terms expired, they were forced to find regular employment. The kinds of jobs varied within sites—some were gas-station attendants, some worked at a printer shop—and men and women frequently performed different kinds of work.

Regresión

1. Usando los datos de la evaluación experimental, estime el efecto causal del programa con una regresión con los ingresos como variable dependiente y la participación en el programa como variable independiente. Muestre su resultado en la columna (1) de su Tabla 1 (ver ejemplo al final del enunciado). ¿Cuál es el efecto estimado del programa?

Suponga ahora, como hizo Lalonde y luego Dehejia & Wahbba, que no dispone de un grupo de control experimental, por lo que debe construir un grupo de comparación no experimental a partir de individuos que no participaron en el programa. Para esto, deberá i) eliminar a todos los individuos del grupo de control experimental y ii) agregar a la base de datos los individuos de la encuesta CPS. Estos individuos serán un primer grupo de comparación no experimental.

2. Estime una regresión análoga a la estimada en la pregunta anterior (con los ingresos como variable dependiente y la participación en el programa como

variable independiente), pero esta vez utilizando el grupo de comparación no experimental. Muestre su resultado en la columna (2) de su Tabla 1 (ver ejemplo al final del enunciado). ¿Cuál es el efecto estimado?, ¿En qué supuesto descansa este análisis?, ¿Le parece un supuesto plausible? Explique.

Propensity Score Matching

3. Construya una tabla (llámela Tabla 2) en la que se comparen las características de individuos del grupo experimental de tratamiento e individuos no tratados identificados en la encuesta CPS. La tabla solicitada se verá como la Tabla 2 que se presenta más abajo. ¿Qué puede concluir a partir de esta tabla?

Tabla 2: Balance in covariates - unmatched data

	Media (Tratados)	Media (Controles)	Diferencia	Error Estándar	Observaciones (Tratados)	Observaciones (Controles)
Age	25.82	33.23	-7.409	0.533	185	15992
Education	10.35	12.03	-1.682	0.150	185	15992
Married	0.19	0.71	-0.522	0.029	185	15992
No degree	0.71	0.30	0.412	0.034	185	15992
Black	0.84	0.07	0.770	0.027	185	15992
Hispanic	0.06	0.07	-0.012	0.018	185	15992
Earnings 1974	2095.57	14016.80	-11921.227	367.155	185	15992
Earnings 1975	1532.06	13650.80	-12118.748	247.777	185	15992

Fuente: Elaboración propia a partir de los datos del estudio.

4. Estime una regresión como la de la pregunta 2., pero esta vez controle en forma lineal por las características observables que incluyó en su Tabla 2. Agregue además, los cuadrados de las variables *edad*, *educación*, y de los ingresos en el año 1974 y 1975. Muestre sus resultados en la columna (3) de su Tabla 1 (ver ejemplo al final del enunciado). ¿Se asemeja su estimación al resultado de la evaluación experimental?
5. Estime un propensity score usando un modelo logit con las variables incluidas en la Tabla 2, y presente un histograma como el de la Figura 1, en el que se describa la distribución del *propensity score* estimado para tratados y no tratados, dejando fuera a individuos con $pscore < 0.05$ o $pscore > 0.95$ (para asegurar soporte común). Interprete la figura.

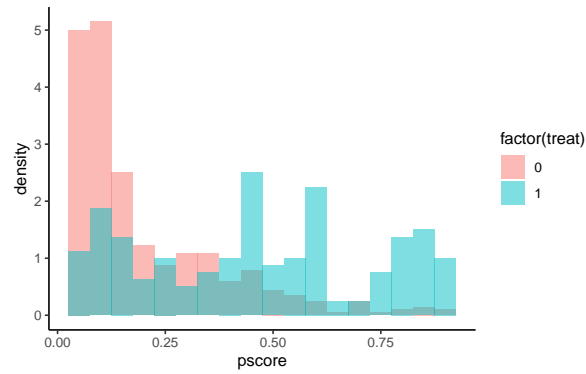


Figura 1: Propensity score estimado, pre-matching

6. Realice un *propensity score matching*, buscando para cada individuo tratado en la muestra experimental, el vecino con el *propensity score* más cercano entre los individuos de la encuesta CPS. Presente un histograma como el de la Figura 2 que muestre la distribución del *propensity score* estimado para los tratados y no tratados emparejados vía *matching*. ¿Qué puede concluir a partir de este análisis? [Para esta pregunta, se recomienda revisar el comando ‘matchit’ de la librería ‘MatchIt’].

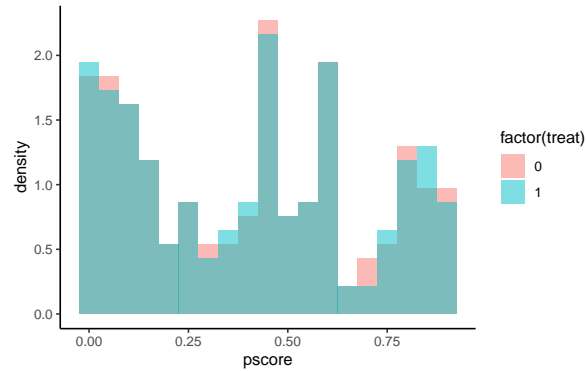


Figura 2: Propensity score estimado, post-matching

7. Construya una tabla análoga a su Tabla 2, en la que se muestre el balance en características de base comparando a tratados y a no-tratados emparejados vía *matching*. Ésta será su Tabla 3 (ver ejemplo más abajo).
8. Estime una regresión como la de la pregunta 2., pero esta vez utilice solo las unidades tratadas y sus respectivos match. Incluya su resultado en la columna (4) de su Tabla 1 (ver ejemplo al final del enunciado). ¿Se asemeja su estimación al resultado de la evaluación experimental?

Tabla 3: Balance in covariates - matched data

	Media (Tratados)	Media (Controles)	Diferencia	Error Estándar	Observaciones (Tratados)	Observaciones (Controles)
Age	25.82	25.69	0.124	0.717	185	185
Education	10.35	10.67	-0.324	0.194	185	185
marr	0.19	0.24	-0.049	0.043	185	185
No degree	0.71	0.70	0.011	0.048	185	185
Black	0.84	0.85	-0.005	0.038	185	185
Hispanic	0.06	0.05	0.005	0.024	185	185
Earnings 1974	2095.57	2023.21	72.361	486.330	185	185
Earnings 1975	1532.06	1367.66	164.394	310.883	185	185

Fuente: Elaboración propia a partir de los datos del estudio.

Inverse Propensity Score Weighting

9. Estime una regresión como la de la pregunta 2., pero ponderando las observaciones por $w_i = D_i + (1 - D_i) \cdot \frac{P(X_i)}{1 - P(X_i)}$ para obtener una estimación del ATT. Muestre su resultado en la columna (5) de su Tabla 1 (ver ejemplo al final del enunciado). ¿Se asemeja su estimación al resultado de la evaluación experimental?

Tabla 1: Estimaciones del efecto causal del programa

	Experimental (1)	Naive Comparison (2)	Linear Controls (3)	PS Matching (4)	PS Weighting (5)
treat	1,794*** (671)	-8,498*** (583)	1,199* (626)	2,517*** (703)	1,570** (759)
Constant	4,555*** (340)	14,847*** (76)	8,719*** (957)	3,832*** (400)	4,779*** (492)
Observations	445	16,177	16,177	370	16,177

Note:

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