



MÉTODOS CUASI-EXPERIMENTALES

PROYECTO FUESE



¿Qué son?

Los métodos cuasiexperimentales son estrategias de evaluación de impacto que buscan aproximarse a la inferencia causal sin asignación aleatoria. Su propósito es responder si los cambios observados en una población pueden atribuirse razonablemente a una intervención, construyendo un contrafactual de forma analítica a partir de reglas de elegibilidad, puntos de corte, datos históricos, cohortes, territorios o variaciones externas.

TABLA DE CONTENIDO

¿Qué son?

¿Qué los distingue?

¿Cuándo usarlos?

Diferencia en Diferencia

Emparejamiento por puntaje de propensión

¿Qué los distingue?

A diferencia de los diseños experimentales, quien evalúa no controla la asignación al tratamiento.

En su lugar, aprovecha reglas de elegibilidad, puntos de corte, datos antes y después, cohortes, territorios, registros administrativos o variaciones externas que permiten construir grupos de comparación plausibles.

Su solidez depende de supuestos verificables y de la transparencia del diseño analítico. No eliminan la necesidad de teoría del cambio: requieren saber qué resultado se espera, en quién, cuándo y por qué. El Banco Mundial y el BID identifican los métodos cuasiexperimentales como base para establecer relaciones causales en evaluaciones de impacto cuando la aleatorización no es posible.

Métodos de esta familia



Las metodologías que destacan en este grupo son: Diferencia en diferencia, series de tiempo interrumpidas, regresión discontinua, y pareamiento. Esta última incluye técnicas como Emparejamiento por Puntaje de Propensión, emparejamiento exacto, Emparejamiento por distancia de Mahalanobis.

En este resumen se incluye información de Diferencia en Diferencia, y Emparejamiento por Puntaje de Propensión.

En la caja de herramientas también se incluye un resumen específico para series de tiempo interrumpidas.

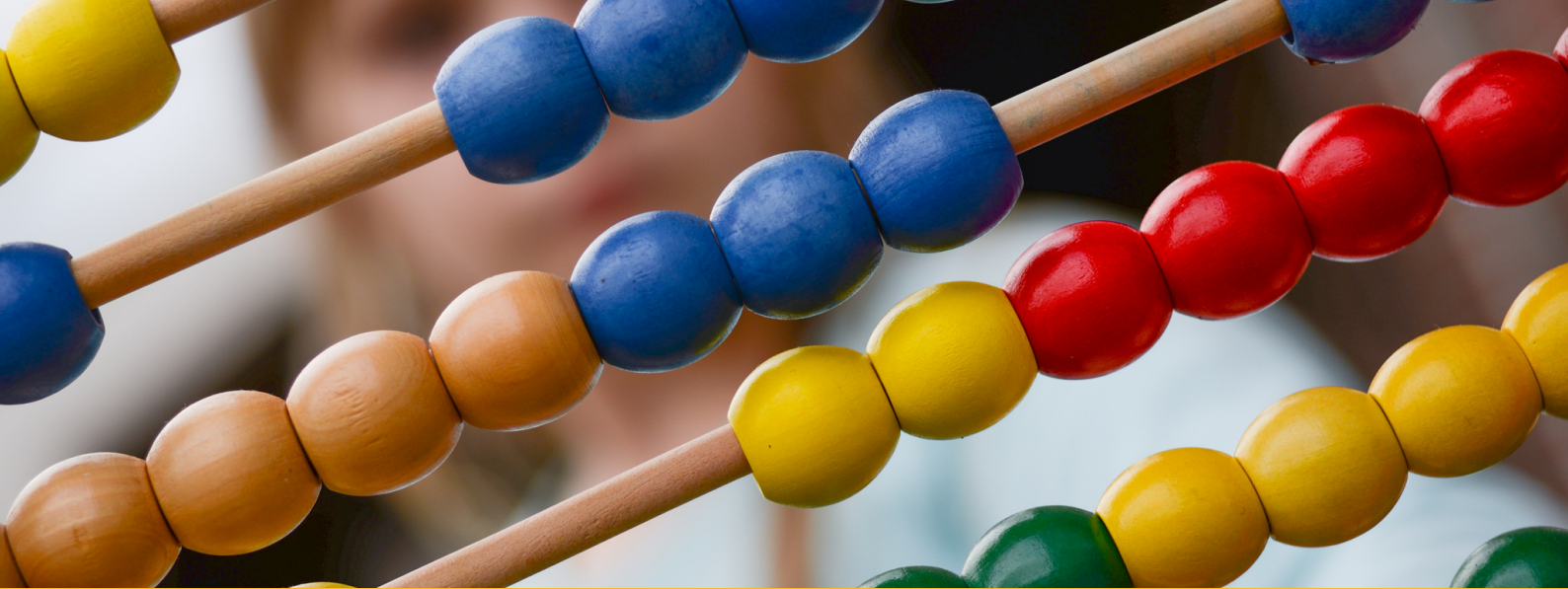
¿Cuándo usarlos?



Estos métodos son recomendables cuando la universidad no puede aleatorizar, pero dispone de registros administrativos, líneas de base, datos comparables, grupos semejantes, series temporales, puntos de corte o cohortes, además de apoyo técnico en análisis cuantitativo. Son especialmente útiles para evaluar tutorías, becas, acompañamiento técnico, formación comunitaria, programas territoriales, políticas de permanencia o proyectos con implementación gradual. Solo es recomendable su aplicación cuando existen ciertas condiciones institucionales que se describen más adelante.

Según el Banco Interamericano de Desarrollo (2017), este grupo incluye métodos que permiten construir comparación plausible sin asignación aleatoria, aprovechando características observables, reglas de asignación, series temporales o variaciones externas.





DIFERENCIA EN DIFERENCIA

Es utilizado por investigadores y científicos de datos cuando la aleatorización no es factible. Este método estima las relaciones causales mediante una doble comparación, de ahí su nombre.

Estima el impacto de una intervención comparando dos cambios primero se calcula la diferencia antes-después para el grupo de tratamiento, luego la diferencia antes-después para el grupo de comparación, y finalmente “la diferencia entre la diferencia” de ambos grupos constituye la estimación del impacto.

¿Cuándo usar esta metodología?

Puede ser útil cuando desea estimar el efecto de una intervención comparando cómo cambia un resultado antes y después entre un grupo que recibió la intervención y otro grupo similar que no la recibió.

Pero no debe usarse solo porque “hay datos antes y después”; requiere condiciones metodológicas fuertes.

Aspectos clave antes de usar Dif-n-Dif

1. Que exista una intervención claramente identificable

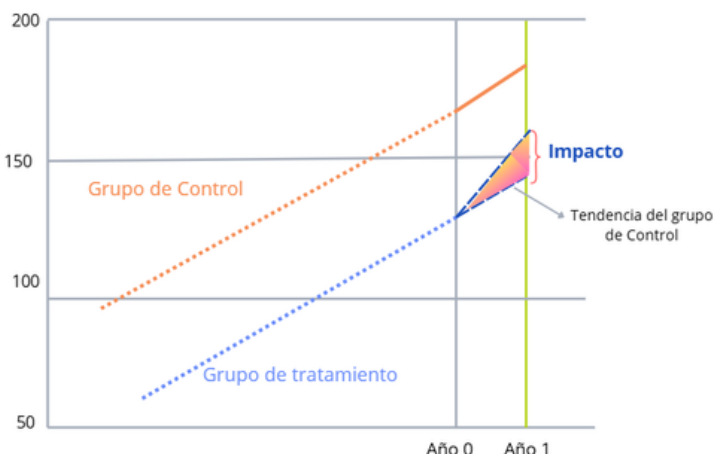
La universidad debe poder delimitar con precisión qué intervención se quiere evaluar, cuándo inició, a quiénes alcanzó y qué resultado se espera modificar. Por ejemplo: una estrategia de acompañamiento académico, un programa de tutorías, una reforma curricular, una intervención de extensión universitaria o una política de becas.

La elección del método de evaluación depende de las características operativas del programa: recursos disponibles, criterios de elegibilidad y plazos de implementación



La metodología se denomina “diferencia en diferencias” porque estima el impacto de una intervención comparando dos cambios: el cambio observado en el grupo de tratamiento antes y después de la intervención, y el cambio observado en el grupo de comparación durante el mismo período. El impacto se obtiene al restar ambas diferencias, por lo que el estimador corresponde literalmente a una diferencia entre diferencias.

Este modelo se basa en la premisa de las tendencias paralelas, es decir que ambos grupos deben seguir tendencias paralelas antes de la intervención.



2. Que haya grupo de tratamiento y grupo de comparación

En una universidad, esto puede traducirse en comparar, por ejemplo, sedes donde se implementó una nueva estrategia de extensión con sedes similares donde aún no se implementó.

3. Disponibilidad de datos

Se requieren datos antes y después de la intervención; no solo se trata de una buena línea base. Lo ideal es contar con al menos dos rondas de medición previo a la intervención, porque eso permite verificar si ambos grupos venían comportándose de manera similar. Esto es esencial para establecer el supuesto de tendencias paralelas, que lo que quiere decir es precisamente que tanto la unidad o grupo de tratamiento como la de control compartían tendencias similares.

4. Que no existan eventos externos que afecten de forma distinta a los grupos

Dif-n-Dif puede fallar si ocurre algo al mismo tiempo que la intervención y afecta solo al grupo de tratamiento o solo al de comparación.

Por ejemplo, si una universidad evalúa una estrategia de tutorías en una sede, pero justo en esa sede también se cambia el equipo docente, se mejora la conectividad, se abre una beca nueva o se modifica la oferta académica, podría atribuirse falsamente a la tutoría un cambio que en realidad se debe a otros factores.

Cualquier factor que afecte desproporcionadamente a uno de los grupos en el mismo momento del tratamiento puede invalidar o sesgar la estimación.

5. Que los indicadores sean estables y comparables en el tiempo

La universidad debe asegurarse de que el resultado medido sea el mismo antes y después. No conviene usar Dif-n-Dif si cambió la forma de medir el indicador.

Por ejemplo, si antes se medía “deserción” de una manera y después se redefine como “interrupción temporal”, la comparación puede ser inválida. Lo mismo ocurre si una sede registra datos de manera distinta a otra.

6. Que la intervención no contamine al grupo de comparación

Debe revisarse si el grupo de comparación realmente no recibió la intervención ni una versión parecida. En universidades esto es frecuente porque las prácticas se difunden informalmente: docentes, extensionistas, sedes o equipos pueden compartir materiales, asesorías o estrategias.

Si el grupo de comparación también recibe parte del tratamiento, el efecto estimado puede reducirse artificialmente.



7. Que exista suficiente tamaño de muestra

DD requiere una cantidad suficiente de casos en ambos grupos y en ambos momentos. Si hay pocos estudiantes, pocas sedes o pocas comunidades, la estimación puede ser muy imprecisa.

En evaluaciones universitarias, esto es especialmente importante cuando se trabaja con proyectos de extensión pequeños, grupos comunitarios reducidos o sedes con baja matrícula.

8. Que se considere la combinación con otros métodos

El Dif-n-Dif puede combinarse con otras técnicas, como emparejamiento, para mejorar la comparabilidad entre grupos.

9. Que se hagan pruebas de validez

Antes de confiar en los resultados, conviene realizar pruebas como:

- Tendencias previas , que responde a la pregunta ¿Los grupos evolucionaban parecido antes de la intervención?
- Placebo con grupo falso que responde a la pregunta: ¿Aparece efecto donde no debería haberlo?
- Placebo con resultado falso, que responde a la pregunta ¿La intervención parece afectar una variable que no debería cambiar?
- Varios grupos de comparación que responde a la pregunta: ¿El efecto se mantiene al usar otros grupos plausibles?

Para que las puntuaciones de propensión estimen correctamente la probabilidad de participación, las características incluidas en su estimación deben estar bien consideradas y ser lo más exhaustivas posible.

Sin embargo, es fundamental no incluir características que puedan haber sido afectadas por el tratamiento. Por este motivo, es recomendable utilizar datos de referencia para estimar las puntuaciones de propensión, si están disponibles.

Una vez seleccionadas todas las covariables relevantes, se realiza una regresión logística o probit y se obtienen las probabilidades predichas.

Emparejamiento por Puntaje de Propensión (PSM)

Es una técnica que se usa para comparar personas, grupos o instituciones que participaron en un programa con otras que no participaron, pero que tenían características similares antes de la intervención.

La idea central es esta: como no siempre podemos asignar aleatoriamente quién recibe un programa y quién no, intentamos construir un grupo de comparación lo más parecido posible al grupo beneficiario. Para eso se calcula una puntuación de propensión, que es la probabilidad de que una persona o unidad participe en el programa según sus características observables, por ejemplo edad, sexo, nivel educativo, zona de residencia, condición socioeconómica u otras variables relevantes.



¿CUÁNDO USAR PSM?

Cuando la universidad tiene datos detallados de personas, comunidades, cursos o proyectos antes de la intervención; cuando la participación no fue aleatoria, pero sí se conoce por qué algunas unidades participaron y otras no; cuando existen suficientes no participantes para construir comparación.

ASPECTOS A TOMAR EN CONSIDERACIÓN

1.Existencia de un grupo de no participantes: aquí debemos preguntarnos: *¿Hay estudiantes, proyectos, sedes, carreras o personas similares que no hayan recibido la intervención?*

El método busca comparar participantes con no participantes parecidos. Si todos recibieron el programa, o si los no participantes son muy distintos, no hay buen grupo de comparación.

2.Calidad de los datos de línea base. La pregunta de este ¿La universidad cuenta con información previa a la intervención?

Deben usarse características observables de línea de base, porque las variables medidas después del programa pueden haber sido afectadas por la intervención y sesgar el resultado (BID, 2017).

3.Variables relevantes para explicar la participación. El PSM funciona mejor cuando se incluyen variables que explican tanto la participación como el resultado. Por ejemplo: rendimiento previo, condición socioeconómica, sede, carrera, edad, sexo, zona de residencia, modalidad de estudio, carga académica, historial de matrícula o apoyos previos.

4.No incluir variables afectadas por el programa. Si el proyecto brinda una asesoría académica, el promedio posterior a la asesoría no debería usarse como variable de emparejamiento, porque ese promedio podría ser efecto de la asesoría misma.

5.Disponibilidad de bases de datos suficientemente grandes. El pareamiento requiere bases de datos amplias. El manual indica que esta técnica usa grandes bases de datos y herramientas estadísticas para construir el mejor grupo de comparación posible con base en características observables.

6.Existencia de “soporte común” o superposición. Si algunos beneficiarios tienen una probabilidad muy alta de participar y no existe ningún no participante parecido, esos casos no pueden compararse adecuadamente. El manual advierte que puede haber falta de rango común o falta de superposición entre tratados y no tratados.

7.Verificación del equilibrio. Después de correr el modelo se compueba que el grupo tratado y el grupo de comparación sean similares en las variables usadas para emparejar. Si no hay equilibrio, debe ajustarse el modelo, cambiar variables o probar otro algoritmo

8.Riesgo de sesgo por variables no observadas. Este es el punto más delicado. El PSM solo controla diferencias observables. Si aspectos como motivación, redes de apoyo, liderazgo docente, compromiso institucional o expectativas personales influyen en participar y también en los resultados, la estimación puede quedar sesgada variables no observadas

9.Capacidad técnica institucional. Aunque los programas estadísticos pueden automatizar parte del procedimiento, la decisión sobre variables, supuestos, equilibrio, interpretación y limitaciones requiere criterio técnico.

10.Interpretación prudente de los resultados. El efecto estimado debe leerse con cautela: “los participantes tuvieron mejores resultados que no participantes comparables”, no necesariamente “el programa explica todo el cambio observado”.

BIBLIOGRAFIA

- Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490), 493-505.
- Baker, A., Callaway, B., Cunningham, S., Goodman-Bacon, A., & Sant'Anna, P. H. C. (2026). Difference-in-differences designs: A practitioner's guide. *Journal of Economic Literature*, 64(2), 498–557. <https://doi.org/10.1257/jel.20251650>
- Bifulco, R., Rubenstein, R., & Sohn, H. (2017). Using synthetic controls to evaluate the effect of unique interventions: The case of Say Yes to Education. *Evaluation Review*. <https://journals.sagepub.com/doi/10.1177/0193841X17742233>
- Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). rdrobust: Software for regression-discontinuity designs. *The Stata Journal*, 17(2), 372-404. <https://journals.sagepub.com/doi/10.1177/1536867X1701700208>
- DIME Wiki. (s. f.). Quasi-experimental methods. World Bank. https://dimewiki.worldbank.org/Quasi-Experimental_Methods
- DIME Wiki. (s. f.). Difference-in-Differences. World Bank. <https://dimewiki.worldbank.org/Difference-in-Differences>
- DIME Wiki. (s. f.). Matching. World Bank. <https://dimewiki.worldbank.org/Matching>
- DIME Wiki. (s. f.). Regression Discontinuity. World Bank. https://dimewiki.worldbank.org/Regression_Discontinuity
- Gertler, P. J., Martínez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2017). *La evaluación de impacto en la práctica* (2.ª ed.). Banco Interamericano de Desarrollo y Banco Mundial. Disponible en: <https://docslib.org/doc/3084979/la-evaluaci%C3%B3n-de-impacto-en-la-pr%C3%A1ctica-segunda-edici%C3%B3n>
- Heinrich, C., Maffioli, A., & Vázquez, G. (2010). A primer for applying propensity-score matching. Inter-American Development Bank. <https://publications.iadb.org/en/primer-applying-propensity-score-matching>