



**UNIVERSIDAD
TORCUATO DI TELLA**

Master in Management + Analytics

Tecnología y Rendimiento Escolar: ¿Cómo afecta el acceso a computadoras en el hogar?

Ezequiel Borrego

Tutor: Pablo Roccatagliata

Junio, 2021

Resumen

Las pruebas Aprender son el dispositivo nacional de evaluación de los aprendizajes de los estudiantes y de sistematización de información acerca de las condiciones en las que ellos se desarrollan en la República Argentina. En base a esta información se busca responder la pregunta: ¿Tener acceso a una computadora en el hogar impacta en el rendimiento escolar? Esta incógnita se responde mediante la utilización de técnicas de inferencia causal, particularmente mediante matching y por propensity score matching. El puntaje de propensión fue calculado mediante regresiones logísticas y algoritmos de boosting a fin de evaluar posibles diferencias. Los resultados de este trabajo pueden ayudar a los responsables de políticas educativas en la asignación de recursos. Los resultados obtenidos sugieren que el impacto de la posesión de una computadora es positivo dentro de los límites del percentil 1 y 99 de la distribución de ATEs estimados. Los resultados varían por técnica, pero se mueven dentro del rango de 0,012 a 0,116 desvíos estándar para lengua y de 0,014 a 0,098 desvíos estándar para matemática en los percentiles mencionados. Además, se presenta un marco de referencia para la evaluación costo beneficio de proyectos educativos basado en (Dhaliwal, Duflo, Glennerster, & Tulloch, 2011). Como chequeo adicional y para contemplar un posible sesgo de selección se realiza una corrección por los missings en la estimación del ATE. En este caso, bajo el supuesto de MAR, consideraremos los missings como parte del mecanismo de selección muestral. Siguiendo la tradición del inverse probability weighting method (IPW) propuesto por (Horvitz & Thompson, 1952) ponderamos por la probabilidad de tener completo el puntaje del examen condicional a los controles (X) bajo el supuesto del cumplimiento del supuesto de MAR. Los resultados obtenidos son similares a los ejercicios anteriores en cuanto al impacto del tratamiento, tanto en los puntajes de lengua como de matemática.

Abstract

Aprender is the national device for evaluating student learning and for systematizing information about the conditions in which they develop in the Argentine Republic. Based on this information, this study is designed to answer a single question: ¿Does having access to a computer at home impact school performance? This question is answered using causal inference techniques, particularly through matching and propensity score matching. Propensity score was calculated with logistic regressions and boosting algorithms to evaluate possible differences. The results of this work can assist educational policy makers in resource allocation. The results obtained suggest that the impact of owning a home computer is positive within the limits of the 1st and 99th percentile of the estimated ATE distribution. Results vary by technique but range from 0,012 to 0,116 standard deviations for language and 0,014 to 0,098 standard deviations for mathematics within the mentioned percentiles. In addition, a reference framework for the cost-benefit evaluation of educational projects is presented based on (Dhaliwal, Duflo, Glennerster, & Tulloch, 2011). As an additional examination and to contemplate the possibility of selection bias, a correction is made for the missings in ATE estimation. In this case, under MAR assumption, we will consider the missings as part of the sample selection mechanism. Following the tradition of the inverse probability weighting method (IPW) proposed by (Horvitz & Thompson, 1952), we weighted by the probability of having the complete test score condition on controls (X) under the assumption of compliance with the MAR assumption. The results obtained are similar to the previous exercises in terms of the impact of the treatment, both in the language and mathematics scores.

Índice

1.	Introducción	5
2.	Marco conceptual y evidencia empírica en educación	6
3.	Metodologías de evaluación de impacto causal	11
3.1	Variables instrumentales	15
3.2	Diseño de regresión discontinua	16
3.3	Diferencias en diferencias	17
3.4	Matching por puntaje de propensión	18
3.5	Desarrollo de técnicas de Matching	18
3.5.1	Matching basado en el puntaje de propensión	20
3.5.2	Matching mediante estrategia de estratificación	22
3.5.3	Matching mediante estrategia de K vecinos más cercanos	23
3.5.4	Estimación del ATE mediante la ponderación del puntaje de propensión	23
3.6	Diferencia con técnicas de Machine Learning	24
3.7	Modelos Causales Gráficos	25
3.7.1	Cadenas	26
3.7.2	Bifurcación	27
3.7.3	Colisión	28
3.7.4	Sesgo de Confusión	28
3.7.5	Sesgo de Selección	29
4.	Datos y Materiales	31
5.	Análisis Descriptivo	32
5.1	Tratamiento Missings	32
5.2	Datos generales de los alumnos	33
5.3	Datos Socioeconómicos	35
5.3	Condiciones habitacionales	37
5.4	Nivel educativo y cultural familiar	38
5.5	Trabajo y tareas hogareñas	39
5.6	Compromiso y conexión con la escuela	40
5.7	Ámbito de la Educación	42
5.8	Resultados en Exámenes	43
5.8.1	Resultados Lengua	44

5.8.2 Resultados Matemáticas	45
6. Estimación del Average Treatment Effect (ATE)	47
6.1 Selección de variables	47
6.2 Técnicas utilizadas	48
6.3 Cálculo del Propensity Score	49
6.4 Controles y Supuestos	49
6.5 Resultados	52
6.6 Corrección del sesgo generado por missings en la variable de respuesta	56
7. Marco de Referencia para Análisis de Efectividad del Costo	58
7.1 Cuantificando el impacto	59
7.2 Cuantificando costos	60
7.3 Ajustes de unidades	61
7.4 Consideraciones Generales	62
8. Conclusiones	64
9. Posibles Extensiones	65
10. Bibliografía	67

1. Introducción

Según los datos recolectados por el observatorio argentino por la educación (Panel de indicadores, 2020), en el año 2018 se registró una tasa de abandono interanual de 0,96%. Al mismo tiempo, de acuerdo con el anuario estadístico educativo 2019 (Ministerio de Educación de la República Argentina, 2020), la tasa de repitencia es de **9,76%**. La Organización para la Cooperación y el Desarrollo Económico (OCDE) lanza anualmente un informe acerca del estado de la educación alrededor del mundo (OECD, 2020). En su último informe correspondiente al año 2020 afirman que en promedio los trabajadores con estudios secundarios ganan un 23% más que quienes no cuentan con estos estudios. En la misma publicación afirma que las diferencias en ingresos se acentúan con la edad entre quienes tienen o no estudios secundarios o terciarios.

La motivación de este estudio es estimar el efecto causal de contar con una computadora en el rendimiento escolar de los alumnos, para así poder evaluar si intervenir con la entrega de una computadora tiene un impacto positivo en su educación. A su vez, en una esfera de acción mayor, la reducción de la repitencia y del abandono escolar tendrían como resultado una sociedad más productiva y menores costos al estado.

2. Marco conceptual y evidencia empírica en educación

Las problemáticas anteriormente mencionadas, la repitencia y el abandono escolar, podrían ser consideradas dos fenómenos independientes. No obstante, el enfoque desarrollado por Jeremy D Finn (Finn J. D., 1989) establece una diferencia sustancial con la visión más tradicional de que el abandono escolar está relacionado a características del individuo o de la institución, y propone dos modelos en los cuales plantea que el abandono es un proceso gradual que puede comenzar incluso desde los primeros años de escolarización.

El primero de los modelos llamado “Autoestima-Frustración” plantea que la baja performance en la escuela lleva a una visión dañada de sí mismo por parte del alumno, la cual a su vez deriva en una oposición del alumno al contexto educativo. Según este modelo esta oposición se manifiesta en comportamientos disruptivos en el aula, la falta a clases y en otro tipo de comportamientos antisociales. Al mismo tiempo, la atención de los adultos responsables por su educación pasa a estar enfocada en controlar su comportamiento y se desatiende su aprendizaje. En una suerte de círculo vicioso, el problema se ve agravado. Estudios más recientes (Branson, Hofmeyr, & Lam, 2014) concuerdan en esta visión del círculo vicioso generado por el mal rendimiento escolar y que un largo período de acumulación deriva en el abandono escolar. Una de las conclusiones del estudio es que la repitencia de grado no es efectiva para impedir el abandono escolar.

Una de las bases que tuvo el modelo de Finn es el aumento del interés de los investigadores educativos en el autoconcepto del alumno en cuanto a su rendimiento escolar, dato que encontraremos entre los que se analizarán para estudiar el rendimiento escolar. Existen estudios que encontraron correlaciones entre la autopercepción escolar y el rendimiento académico (Byrne, 1984), pero no lograron probar causalidad.

Por otro lado, el otro modelo propuesto por Finn es el de “Participación – Identificación”. Este está basado fuertemente en un famoso experimento conocido como “Perry Preschool Project” (Berrueta-Clement, 1984) en el que se experimentó y documentó los efectos de atender a la escuela preescolar para 5 camadas de chicos desde los 4 años hasta los 19. Los grupos de alumnos fueron divididos en tratamiento y control en todas las ‘olas’ de chicos, y se aseguraron de que no haya diferencias estadísticamente significativas entre los grupos de control y tratamiento en cada una de ellas en aspectos como ratio de hombres y mujeres, nivel socioeconómico, resultados de prueba de IQ, empleo de los padres, y otros aspectos de la vida familiar de los alumnos.

Los resultados fueron muy alentadores en cuanto a los efectos positivos que tiene acceder al preescolar, y en particular, una de las razones planteadas por el autor es la relación social que se genera entre las personas y el ambiente, en este caso el colegio, que son observados como factores que evitan el desarrollo de actitudes delincuenciales, mientras que vínculos débiles con el establecimiento y el resto de los alumnos hacen que esto sea más probable.

De allí que el modelo planteado por Finn se basa en reducir la alienación del alumno con el establecimiento educativo, planteando que la participación voluntaria de los estudiantes, los claros objetivos educativos, un tamaño relativamente pequeño de la escuela, la participación en la toma de decisiones por parte de los alumnos, las relaciones entre los alumnos y que el trabajo sea significativo para el alumno contribuyen a fortalecer los vínculos entre el alumno y la escuela. Las actividades extracurriculares también son destacadas como factores importantes de vinculación en los casos de alumnos con resultados académicos pobres.

El autor destaca la importancia de descubrir la “no participación” lo antes posible, ya que es un proceso largo que se va desarrollando por muchos años. Por mucho tiempo no se exploró este fenómeno debido a que los datos existentes de los alumnos que abandonan el colegio eran solamente del mismo año en el que abandonan o del anterior. Hasta que en el año 1978 se realizó un estudio longitudinal (Lloyd, 1978), en el que se dio seguimiento a la evolución de 1500 alumnos de sexto grado hasta el final del secundario. Sumado a esto, se analizó el desempeño escolar de estos chicos en el tercer grado una vez sabido el resultado final de si abandonan o no el secundario. Los resultados fueron novedosos ya que se encontraron diferencias significativas en el desempeño escolar de los futuros desertores y los que completaron la escuela. Este modelo sigue vigente en literatura académica más reciente (Rumberger & Rotermund, 2012), en el que se mantiene la hipótesis de que el abandono escolar es un proceso y no un evento, que no hay un solo factor que lo cause y que dos de los indicadores más consistentes son el desempeño escolar en los primeros años y la participación académica y social de los alumnos con la escuela. El mismo estudio indica también que la intervención temprana es recomendable.

Los alumnos que abandonan difieren en (Rumberger R. W., 1987) y para cada uno de ellos recomienda la utilización de un enfoque particular para prevenirlo. En algunos casos se expresan problemas con la escuela en sí misma, en otros se enuncian razones externas a ella, como la necesidad de trabajar o el embarazo adolescente. Se destaca también que además del apoyo académico que pueda brindar la institución, también es importante satisfacer la necesidad psicológica de que hay alguien interesado por ellos a nivel individual, lo cual se puede lograr con el asesoramiento de un psicopedagogo.

Rumberger expresa con énfasis que es crucial identificar apropiadamente a los alumnos más predispuestos a abandonar la escuela y que mientras antes se los logre identificar, mayor será la probabilidad de evitar un desenlace de abandono de su educación.

En resumen, lo expuesto por los mencionados autores enfatiza en estudiar el proceso de largo plazo que culmina en el abandono escolar más que en “características estructurales” como el estatus socioeconómico. Si bien Finn plantea dos modelos como posibles, coincide en que ambos identifican la pobre performance escolar como un importante antecedente al abandono escolar (Finn J. D., 1989). Desde un punto de vista de autoestima, los malos resultados impactan en como el alumno se auto percibe. Desde la perspectiva del modelo de participación, los malos resultados son un claro indicador del distanciamiento entre el alumno y sus obligaciones escolares.

En 2005 Finn y Gerber, basados en las teorías enunciadas por Finn anteriormente, estudiaron el efecto de atender a clases pequeñas de menos de 20 alumnos en los primeros años escolares en la tasa de graduación ¹. Encontraron que quienes atendían a clases pequeñas por 3 o 4 años lograron incrementar significativamente su tasa de graduación, especialmente para los alumnos de familias de clase social baja. Lo cual reafirma las teorías de Finn respecto al carácter progresivo del abandono escolar y cómo interviniendo tempranamente se puede prevenir este fenómeno.

El estudio de la inferencia causal en el ámbito educativo ha crecido en popularidad en los últimos años (Gustafsson, 2008), una de las razones de este auge es el lanzamiento del programa PISA (Programme for International Student Assessment) por la OCDE (Organización para la Cooperación y el Desarrollo Económicos) en el año 2000, a través del cual se puso énfasis en la evaluación de la calidad de los servicios educativos de los distintos países miembro. Esto generó una infraestructura de datos con la que no se contaba anteriormente, y que está disponible para la comunidad investigativa.

Cordero y Cristóbal (Cordero & Cristóbal, 2017) realizan un resumen de lo más reciente en materia de investigación educativa desde un punto de vista econométrico. Destacan el uso mayoritario de las bases de datos de los exámenes PISA, los cuales son de una naturaleza muy similar a los que contamos con las pruebas ‘Aprender’ de alcance nacional y explican los diferentes tipos de enfoques que se le puede dar a este tipo de análisis.

¹ Finn & Gerber, 2005

El primero de estos casos es una publicación del Banco Interamericano de Desarrollo (Cristia, 2012) en el marco del programa “One Laptop per Child” (OLPC) que apunta a mejorar el aprendizaje en las regiones más pobres del mundo proveyendo computadoras portátiles a niños para su uso en el hogar y en la escuela. El programa se implementó en 36 países y más de dos millones de computadoras fueron entregadas. Se destaca que cada computadora tuvo un costo aproximado de 200 dólares mientras que la inversión por alumno en los países de bajos ingresos fue tan solo de 48 dólares. En el caso de los países de medianos ingresos se estima que se invierten 555 dólares por alumno. Aún con esos datos de inversión relativa en computadora y educación había poca evidencia sólida sobre la efectividad del programa. El estudio en cuestión se centró en 319 escuelas rurales de Perú, en el cual se recopilaban datos luego de 15 meses de la implementación del programa para evaluar si el aumento del acceso a computadoras logró un incremento en el capital humano. Principalmente se evaluaron los resultados académicos en matemática y lengua, las habilidades cognitivas, los resultados de exámenes de fluencia verbal y de programación. La investigación se basó en una prueba controlado aleatorizada a nivel escuela con 209 escuelas con tratamiento y 110 escuelas control.

Los resultados del estudio indican un mayor uso de la computadora en la escuela, pasando de un 26% a un 82%, un mayor uso de la computadora en el hogar, el cual pasó de 4% a 42% y una aptitud general en la operación de las computadoras. Respecto a los resultados académicos no se encontraron evidencias que el programa mejore el aprendizaje en matemáticas o lengua, estimando el efecto en ambos en 0.003 desvíos estándar con un error estándar de 0.055. Sí se encontraron resultados positivos en los exámenes cognitivos, fluencia verbal y programación.

Otro importante estudio relacionado es el de Joshua Angrist acerca del uso de computadoras para la enseñanza en Israel². El uso de computadoras para la educación se puede encuadrar en dos grandes patrones: el aprendizaje de competencias de computación (CST: Computer Skill Training), las cuales se centran en enseñar a los alumnos en *cómo* utilizar las computadoras; y en la instrucción apoyada por computadoras (CAI: Computer Aided Instruction), que *utiliza* las computadoras para la enseñanza de temas que pueden o no estar relacionados con la tecnología.

La utilidad de CST ya era apreciada como de gran utilidad, pero el rol de CAI todavía se encontraba en duda, ya que había un interés creciente en el uso de computadoras para la educación, pero no había consenso en su efectividad. El estudio se basó en un programa israelí llamado “Tomorrow-98” en el cual

² Angrist & Lavy, 2002

se buscó la computarización del sistema educativo. El programa consistió en invertir significativamente en entrenar a los profesores en integrar las computadoras en la enseñanza y en equipar a las escuelas con hardware y software moderno. Durante cinco años se equiparon escuelas y se entrenaron profesores para cumplir con el programa en el cual municipalidades y autoridades regionales aplicaban para recibir los fondos. Se obtuvieron datos de 200 escuelas de las cuales 122 aplicaron para el programa. En base a la técnica de variables instrumentales se obtiene un efecto estimado del uso de computadoras para asistir en la enseñanza, el cual fue negativo en los resultados en matemática para los chicos de cuarto grado. Para las otras materias los resultados no fueron significativos, aunque en su mayoría la tendencia era negativa. La reflexión fue que CAI no es mejor que la enseñanza tradicional y que en particular podría hasta ser menos efectiva que otros métodos de enseñanza. Se resaltó el alto nivel de inversión que requirió el programa y se concluyó que probablemente hubiera sido mejor utilizar el dinero en otro tipo de programas.

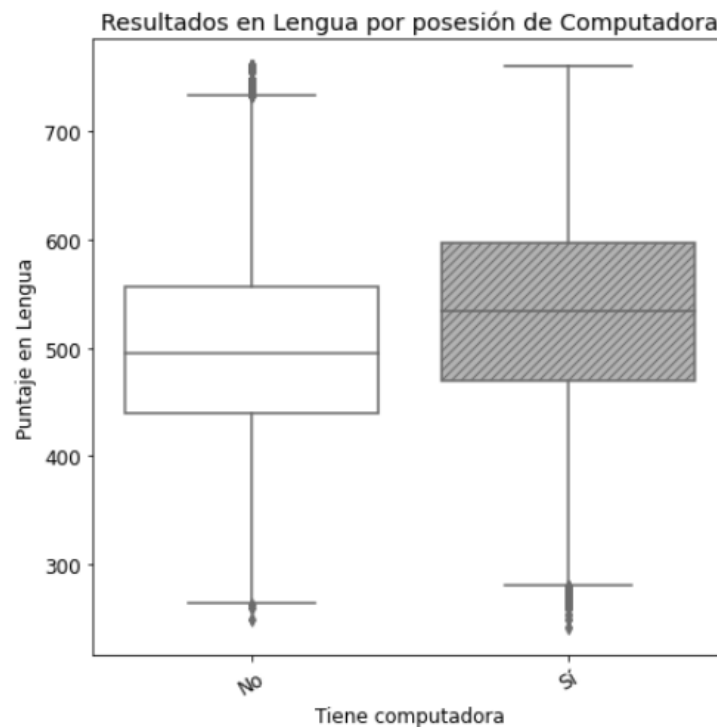
(Fairlie & Robinson, 2013) es nuevamente un experimento controlado que involucró a 1.123 alumnos de sexto a decimo grado de 15 escuelas de California, Estados Unidos. En los Estados Unidos se invierten más de 5 mil millones de dólares anuales en computadoras y tecnología de información para las escuelas, pero se resalta que aproximadamente cerca de nueve millones de niños entre 10 y 17, que componen un 27% del total, no tienen acceso a una computadora en su casa. Se destaca en el estudio que en una investigación previa de experimento controlado se encontraron pequeños efectos positivos en el rendimiento académico de alumnos universitarios pero que no alcanzaba a alumnos de escuelas. Es por ello por lo que se realizó nuevamente un experimento controlado con entrega aleatoria de computadoras a alumnos de escolares. La mitad de los 1.123 recibió computadoras, mientras que la otra no. Cada computadora costó entre 400 y 500 dólares y se recopilaron la mayoría de los resultados 6 meses después de la entrega para así disminuir la posibilidad de que el comportamiento del estudiante haya cambiado por otra razón que no sean las computadoras. Se detectaron grandes impactos en la cantidad de horas de uso de computadoras y en el acceso a internet para el grupo con tratamiento. No se hallaron efectos en el rendimiento académico, por lo que se concluyó que es poco probable que la posesión de computadoras por sí solas tengan impacto en el corto plazo en los resultados escolares de chicos de familias de bajos ingresos.

Teniendo en cuenta los resultados de estas experiencias, procederemos a realizar un estudio de impacto para los alumnos de sexto grado de primaria de la República Argentina.

3. Metodologías de evaluación de impacto causal

El objetivo es identificar el efecto causal de aplicar un tratamiento a un individuo. En este caso, cómo impacta en el desempeño escolar de un alumno el hecho de tener una computadora. En una primera instancia, se puede analizar la distribución de puntajes en un examen para los estudiantes que tienen o no computadora, lo cual será nuestra variable tratamiento.

Figura 1: Distribución de puntajes en Lengua por posesión de computadora



Como se observa en la figura 1, el primer cuartil, la mediana y el tercer cuartil son superiores a sus contrapartes en el grupo que no tiene computadora. Sin embargo, esto no significa que **el hecho de tener computadora** sea el causante de las mejores notas. Lo único que podemos observar es que a los alumnos que tienen computadora pareciera irles mejor. Se podría asumir que la causa de que a los alumnos que tienen computadora les vaya mejor es debido a que tienen mayor proporción de clase media y alta, mientras que los alumnos sin computadoras son en mayor proporción de clase baja. Esto genera el interrogante de si el hecho de que le vaya mejor a un grupo es debido a la computadora o debido a la clase social.

Realizar preguntas del estilo, ¿qué pasaría con el rendimiento escolar si el alumno tuviera una computadora?, es clasificado por Andrew Gelman (Gelman, 2011) como inferencia causal hacia adelante. En este tipo de preguntas, se seleccionan unos potenciales tratamientos para ser estudiados y así conocer “los efectos de las causas”. Este tipo de preguntas no son simples de responder ya que implican poder resolver un gran desafío: Utilizar estudios observacionales o semi experimentales que contienen información faltante.

El hecho de tener información faltante es inherente a los efectos causales. Si se ha tomado un camino en algún punto (tratamiento), solo podrá ser observado el resultado de ese tratamiento, quedando siempre faltante el resultado de haber hecho otra cosa en ese mismo instante. Es por ello por lo que es conveniente establecer el marco teórico sobre el que se operará, el cual es el esquema de resultados potenciales de Rubin (Rubin, 2005). Buscamos determinar si tener una computadora impacta en el rendimiento escolar. Por lo tanto, crearemos un ejemplo ficticio de un alumno al cual llamaremos Juan. Juan no tiene computadora, pero estamos parados en un momento del tiempo en el que podríamos darle una y luego observar que calificaciones obtiene en los exámenes de lengua y matemáticas.

Se denominará el hecho de contar o no con una computadora como **intervención**. La notación que identificará el resultado de Juan será Y_{Juan} .

El tratamiento para Juan será:

$$T_{\text{Juan}} = \begin{cases} 1 & \text{si Juan recibe una computadora} \\ 0 & \text{si no la recibe} \end{cases}$$

El problema fundamental del análisis causal es que no podemos observar para una misma persona el resultado con o sin tratamiento. Cada vez que se opta por una opción se pierde la posibilidad de observar el resultado alternativo. Dicho de otra forma, sólo podremos observar uno de estos dos resultados: $Y_{\text{Juan}}(1)$ o $Y_{\text{Juan}}(0)$.

Imagen 1: Bifurcación de caminos (Alves, 2020)



Solo uno de los caminos podrá tomarse a la vez, una vez elegido uno perderemos la posibilidad de observar que hubiera pasado de tomar el alternativo. Esta es la principal complejidad de este tipo de análisis. Si hipotéticamente pudiéramos tomar ambos caminos, podríamos definir el efecto causal como:

$$Y_{1,Juan} - Y_{0,Juan}$$

Una opción con la que contamos es comparar a dos personas, Juan y Sofía, que hayan tenido distintos tratamientos. Sin embargo, casi instintivamente nos damos cuenta de que esto no es del todo correcto. Juan y Sofía potencialmente pueden ser personas muy diferentes, lo cual nos llevaría a un sesgo de selección el cual podemos plasmar en la siguiente ecuación (Angris & Pischke, 2015):

$$Y_{Juan} - Y_{Sofía} = Y_{1,Juan} - Y_{0,Sofía} = Y_{1,Juan} - Y_{0,Juan} + \{Y_{0,Juan} - Y_{0,Sofía}\}$$

En la ecuación hacemos la comparación entre las calificaciones de Juan que recibió el tratamiento y Sofía que no lo recibió, con la correspondiente notación previo al nombre. Al sumar y restar a esta comparación los resultados de Juan en el universo en que no recibió el tratamiento, podemos identificar dos sucesos interesantes: El primero es $Y_{1,Juan} - Y_{0,Juan}$, que es el efecto causal de entregarle una computadora a Juan. El segundo es $Y_{0,Juan} - Y_{0,Sofía}$, que es la diferencia entre ambos estudiantes antes de hacerles

ningún tratamiento, cuyo resultado indicará si Juan tenía mejor o peor rendimiento que Sofía en este tipo de exámenes. En el contexto de nuestro estudio este término es el **sesgo de selección**.

Nuestro interés está en eliminar el sesgo de selección y en calcular el efecto causal promedio también conocido como ATE (Average Treatment Effect). Para lograrlo partiremos de la siguiente ecuación:

$$Avg_n[Y_{1i} - Y_{0i}] = \frac{1}{n} \sum_{i=1}^n [Y_{1i} - Y_{0i}] = \frac{1}{n} \sum_{i=1}^n Y_{1i} - \frac{1}{n} \sum_{i=1}^n Y_{0i}$$

La fórmula describe una comparación entre cada miembro de la población teniendo y no teniendo el tratamiento. Es decir, compara los universos donde a todos se les dio intervención con el que nadie la recibió. La sumatoria resume de forma elegante lo que sería el efecto causal individual de comparar $Y_{1,Juan} - Y_{0,Juan}$ para cada una de las personas i que componen el universo de n personas.

El análisis del efecto causal promedio (ATE) comienza comparando el resultado promedio de los que sí tuvieron tratamiento contra el resultado promedio de los que no lo tuvieron. Esto se puede escribir con la ayuda de una variable *dummy*, T_i , que toma el valor 1 si el tratamiento fue aplicado, o 0 si el tratamiento no fue aplicado.

$Avg_n[Y_i | T_i = 1]$ sería el resultado promedio de quienes recibieron el tratamiento y $Avg_n[Y_i | T_i = 0]$ el resultado promedio de quienes no lo recibieron. Al ser ambos promedios condicionales al tratamiento que recibieron, el resultado de $Avg_n[Y_i | T_i = 1]$ es el equivalente al promedio de Y_{1i} , lo cual no contiene información de Y_{0i} . Lo mismo ocurre en viceversa para el resultado con el condicional alternativo.

$$\begin{aligned} \text{Diferencia en promedios grupales} &= Avg_n[Y_i | T_i = 1] - Avg_n[Y_i | T_i = 0] \\ &= Avg_n[Y_{1i} | T_i = 1] - Avg_n[Y_{0i} | T_{0i} = 0] \end{aligned}$$

Esto dista de ser lo que nos gustaría averiguar que es:

$$Avg_n[Y_{1i} - Y_{0i}]$$

Que como se ha dicho anteriormente sería comparar el Y_{1i} de todos con el Y_{0i} también de todos. La manera de acercarnos a ello es mediante la siguiente igualdad:

$$Y_{1i} = Y_{0i} + \kappa$$

Es decir, definimos el resultado de una persona con tratamiento como su resultado sin tratamiento sumado a una constante que representa al efecto causal del tratamiento para el individuo. En el caso de una población, la constante equivaldría al efecto causal promedio.

Si retornamos a la ecuación de diferencia en promedios grupales y sustituimos con la fórmula anterior tendríamos:

$$\begin{aligned} Avg_n[Y_{1i} | T_i = 1] - Avg_n[Y_{0i} | T_{0i} = 0] &= \{\kappa + Avg_n[Y_{0i} | T_{0i} = 1]\} - Avg_n[Y_{0i} | T_{0i} = 0] \\ &= \kappa + \{Avg_n[Y_{0i} | T_{0i} = 1] - Avg_n[Y_{0i} | T_{0i} = 0]\} \end{aligned}$$

Al igual que en el ejemplo entre Juan y Sofía, la diferencia entre dos grupos está definida por dos partes:

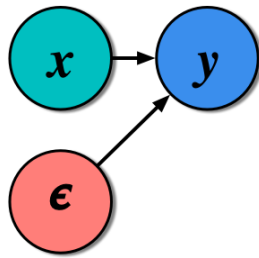
$$\text{Diferencia entre medias grupales} = \text{efecto causal promedio} + \text{Sesgo de selección}$$

Si se lograra aislar el sesgo de selección, podríamos obtener el ATE y así determinar si los tratamientos son realmente efectivos. En la disciplina se destacan cuatro metodologías diferentes para estimar el ATE, las cuales se describirán brevemente.

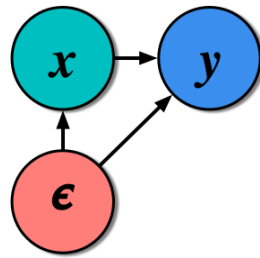
3.1 Variables instrumentales

Es un método que utiliza una variable externa, la cual se llama *instrumento*, para corregir endogeneidad en las variables observacionales. El requerimiento es que el instrumento sea exógeno a las otras variables, lo cual requiere conocimiento del sistema para determinar si hay variables exógenas. Esta variable exógena tiene que correlacionar con la X endógena pero no con Y.

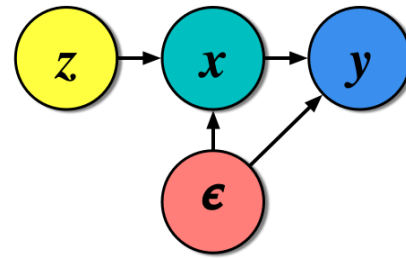
Imagen 2: Tres escenarios comunes de regresión (Mohamed, 2021)



Typical regression



Errors in variables



Instrumental variables regression

La más común aplicación de esta metodología es en los cuasi experimentos, que suelen tener cumplimiento imperfecto, inscripción voluntaria o cobertura universal.³ En los programas sociales del mundo real, es poco realista pensar que el administrador del programa será capaz de asegurar un cumplimiento pleno de la asignación del grupo. La técnica de variables instrumentales sirve para evaluar este tipo de programas en el que no todos los asignados terminan aceptando el rol que se les asigna.

Un ejemplo de un caso posible de aplicación: Comparación de resultados de escuelas a las que se accede mediante una lotería. El ganador de la lotería puede rechazar el cupo a esa escuela y atender a otra. En estos casos existe una diferencia entre el impacto en la *intención de tratar* y *tratamiento en los tratados*. Siendo los primeros los asignados al tratamiento y los segundos los que efectivamente lo aceptan. La variable instrumental en este caso es la lotería que tiene un impacto en la variable X que es recibir el tratamiento, y es a través de esa variable X que indirectamente afecta al resultado Y.

3.2 Diseño de regresión discontinua

Es un método de evaluación de impacto que se puede utilizar en casos donde hay un índice de elegibilidad continuo con un umbral definido que divide a los elegidos de los no seleccionados.⁴

En el área próxima al umbral de selección, la separación para tratamiento resulta aleatoria al no haber diferencias observables entre quienes estén de un lado u otro del umbral. Sin embargo, esto no aplica

³ Gertler, Martínez, Premand, Rawlings, & Vermeersch, 2017

⁴ Gertler, Martínez, Premand, Rawlings, & Vermeersch, 2017

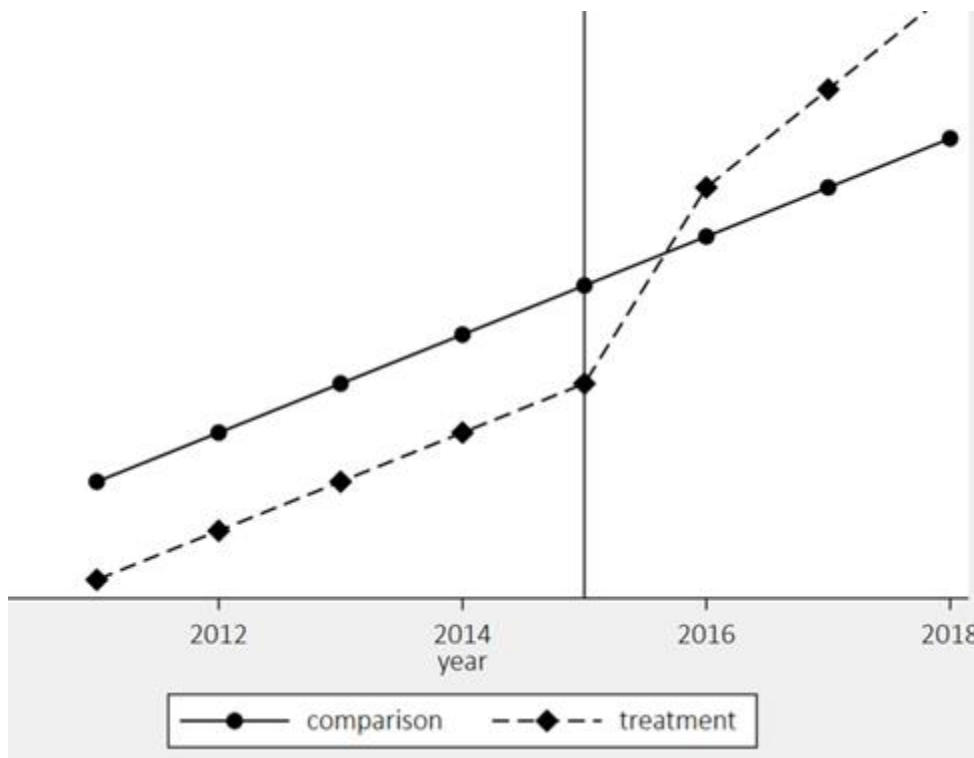
para los individuos que se encuentran más alejados del umbral ya que fuera de la cercanía del punto de corte no se asemeja a un experimento aleatorio y la validez de la comparación disminuye.

Esta técnica es muy útil para la evaluación de políticas educativas basadas en reglas como puede ser una beca o admisión a un establecimiento educativo. (Cordero & Cristóbal, 2017)

3.3 Diferencias en diferencias

Es un método con similitudes al anterior, pero con la diferencia en que se compara el antes y después de una unidad con tratamiento respecto al cambio en el tiempo de otra unidad sin tratamiento. Es decir, se compara la diferencia en los resultados antes-después para el grupo con tratamiento con los resultados antes-después para el grupo de control. (Gertler, Martínez, Premand, Rawlings, & Vermeersch, 2017) De esta manera se controlan los factores que son constantes a lo largo del tiempo cuando se evalúa el antes y el después, y también los factores externos cuando se comparan los grupos de control y tratamiento.

Imagen 3: Diferencias en diferencias (Hillman, 2021)



3.4 Matching por puntaje de propensión

Es un método que permite matchear unidades con tratamiento con unidades sin tratamiento sin la necesidad de que todas las variables de control observables sean iguales. Para cada unidad se calcula la probabilidad de que ésta reciba tratamiento, y posteriormente seorean las unidades del grupo control con las de tratamiento que tengan los valores de propensión más similares. Su punto débil es que es tan bueno como las características que se utilizan para el matching. Si existieran variables no observables que influyen fuertemente en la chance de recibir o no tratamiento, y que también influya en el resultado, entonces las estimaciones estarían sesgadas. (Gertler, Martínez, Premand, Rawlings, & Vermeersch, 2017)

Esta técnica es aplicada usualmente cuando se cuenta con una buena cobertura del espacio de predictores, que serían las variables observacionales, como para poder encontrar observaciones similares en ambos grupos de comparación. Se utilizará esta técnica para estimar el impacto de contar con una computadora para los alumnos de sexto grado de Argentina a partir de los datos de las pruebas Aprender.

Las técnicas de matching y propensity score matching son las que se utilizarán en este estudio y serán desarrolladas en la siguiente sección

3.5 Desarrollo de técnicas de Matching

El siguiente desarrollo se basa en (Galiani, 2021).

Asumiendo que la asignación de tratamiento fuera enteramente aleatoria en la población, los resultados potenciales serían independientes del estado de tratamiento:

$$Y_{(1)}, Y_{(0)} \perp D$$

De esta manera la información faltante, el resultado alternativo para cada observación, no generaría problemas ya que

$$E\{Y_i(0)|D_i = 0\} = E\{Y_i(0)|D_i = 1\} = E\{Y_i(0)\}$$

$$E\{Y_i(1)|D_i = 0\} = E\{Y_i(1)|D_i = 1\} = E\{Y_i(1)\}$$

Entonces, el ATE podría expresarse como:

$$\begin{aligned}
 \delta &= E\{\Delta_i | D_i = 1\} \\
 &= E\{Y_i(1) - Y_i(0) | D_i = 1\} \\
 &= E\{Y_i(1) | D_i = 1\} - E\{Y_i(0) | D_i = 1\} \\
 &= E\{Y_i(1) | D_i = 1\} - E\{Y_i(0) | D_i = 0\} \\
 &= E\{Y_i | D_i = 1\} - E\{Y_i | D_i = 0\}
 \end{aligned}$$

La aleatoriedad asegura que el sesgo de selección de la muestra sea 0:

$$E\{Y_i | D_i = 1\} - E\{Y_i | D_i = 0\} = 0$$

Lo que esto significa es que la información faltante para una observación que tuvo o no tratamiento no es un impedimento ya que la información faltante falta de manera aleatoria, por lo que no habría diferencias significativas entre quienes recibieron tratamiento y quienes no lo recibieron.

Si la asignación de tratamiento aleatoria no es posible entonces es necesario utilizar otro tipo de técnicas, previo a esto es necesario definir el concepto de que no hayan “factores de confusión”. Esto significa que, dentro de cada una de las celdas que componen la matriz de variables observacionales, el tratamiento debe ser aleatorio. En otras palabras, que la selección del tratamiento dependa únicamente de los observables X.

Esto se expresa matemáticamente como:

$$Y_{(1)}, Y_{(0)} \perp D | X$$

Entonces si asumimos que no hay factores de confusión:

$$\begin{aligned}
 E\{Y_i(0) | D_i = 0, X\} &= E\{Y_i(0) | D_i = 1, X\} = E\{Y_i(0) | X\} \\
 E\{Y_i(1) | D_i = 0, X\} &= E\{Y_i(1) | D_i = 1, X\} = E\{Y_i(1) | X\}
 \end{aligned}$$

Esto es similar a lo que expresamos anteriormente cuando la asignación era aleatoria, excepto que esta vez la ausencia de factores de confusión es siempre dentro de una misma celda de la matriz X.

Utilizando las expresiones anteriores podemos definir para cada celda de X:

$$\begin{aligned}
 \delta_x &= E\{\Delta_i|X\} \\
 &= E\{Y_i(1) - Y_i(0)|X\} \\
 &= E\{Y_i(1)|X\} - E\{Y_i(0)|X\} \\
 &= E\{Y_i(1)|D_i = 1, X\} - E\{Y_i(0)|D_i = 0, X\} \\
 &= E\{Y_i|D_i = 1, X\} - E\{Y_i|D_i = 0, X\}
 \end{aligned}$$

Mediante la ley de probabilidad total, el efecto promedio de tratamiento en los tratados está dado por:

$$\begin{aligned}
 \delta &= E\{\Delta_i|D_i = 1\} \\
 &= E\{E\{\Delta_i|D_i = 1, X\}|D_i = 1\} \\
 &= E\{E\{Y_i|D_i = 1, X\} - E\{Y_i|D_i = 0, X\}|D_i = 1\} \\
 &= E\{\delta_x|D_i = 1\}
 \end{aligned}$$

Esto implica que podemos dividir los datos en celdas definidas por cada valor particular de X, luego para cada una de ellas calcular la diferencia entre los resultados de los tratamiento y control y finalmente promediar estas diferencias.

En este punto es importante mencionar el problema conocido como maldición de la dimensionalidad. Este hace referencia a que al aumentar la cantidad de dimensiones la cantidad de celdas posibles aumenta exponencialmente, por lo que es posible encontrarse con celdas que solo contengan unidades de tratamiento o control.

Rosenbaum y Rubin proponen una estrategia basada en el puntaje de propensión⁵ que permite disminuir el problema de la dimensionalidad.

3.5.1 Matching basado en el puntaje de propensión

El puntaje de propensión está definido como la probabilidad condicional de recibir tratamiento dados los controles.

$$p(x) = Pr\{D = 1|X\} = E\{D|X\}$$

⁵ Rosenbaum & Rubin, The central role of the propensity score in observational studies for causal effect, 1983

El puntaje de propensión cuenta con dos propiedades importantes:

1. Balancea las variables observacionales dado un puntaje de propensión:

$$D \perp X | p(x)$$

2. La asignación de tratamiento no tiene factores de confusión dado un puntaje de propensión:

$$Y_{(1)}, Y_{(0)} \perp D | p(x)$$

Utilizando estas propiedades podemos matchear casos de tratamiento y control mediante el puntaje de propensión sin necesidad de utilizar las variables observables X .

$$\begin{aligned} E\{Y_i(0)|D_i = 0, p(X_i)\} &= E\{Y_i(0)|D_i = 1, p(X_i)\} = E\{Y_i(0)|p(X_i)\} \\ E\{Y_i(1)|D_i = 0, p(X_i)\} &= E\{Y_i(1)|D_i = 1, p(X_i)\} = E\{Y_i(1)|p(X_i)\} \end{aligned}$$

Como se hizo anteriormente para cada celda de la matriz X podemos definir el ATE (efecto promedio de tratamiento) mediante el puntaje de propensión:

$$\begin{aligned} \delta_x &= E\{\Delta_i | p(X_i)\} \\ &= E\{Y_i(1) - Y_i(0) | p(X_i)\} \\ &= E\{Y_i(1) | p(X_i)\} - E\{Y_i(0) | p(X_i)\} \\ &= E\{Y_i(1) | D_i = 1, p(X_i)\} - E\{Y_i(0) | D_i = 0, p(X_i)\} \\ &= E\{Y_i | D_i = 1, p(X_i)\} - E\{Y_i | D_i = 0, p(X_i)\} \end{aligned}$$

Aplicando nuevamente la ley de probabilidad total:

$$\begin{aligned} \delta &= E\{\Delta_i | D_i = 1\} \\ &= E\{E\{\Delta_i | D_i = 1, p(X_i)\} | D_i = 1\} \\ &= E\{E\{Y_i(1) | D_i = 1, p(X_i)\} - E\{Y_i(0) | D_i = 0, p(X_i)\} | D_i = 1\} \\ &= E\{\delta_x | D_i = 1\} \end{aligned}$$

Como se mencionó en el caso que utilizamos las variables observables X , podemos dividir los datos en celdas definidas por cada valor de propensión y luego para cada uno de ellos calcular la diferencia entre los resultados de los tratamiento y control y finalmente promediar estas diferencias.

Para utilizar el puntaje de propensión como estrategia para estimar el efecto causal de un tratamiento se requieren solamente de dos pasos:

1. Estimar el puntaje de propensión: No conocemos el valor real de propensión ya que es desconocido, por lo que solo podemos estimarlo.

2. Estimar el efecto promedio de tratamiento (ATE por sus siglas en inglés): Esto se logra pareando casos de tratamiento y control con el mismo puntaje de propensión, después calcular el ATE para cada valor del puntaje de propensión y finalmente obtener el efecto promedio.

Al igual que con matching, no es factible encontrar observaciones con exactamente el mismo valor de propensión. Afortunadamente hay más de una alternativa posible para lograr el pareo, entre ellas la estratificación del puntaje de propensión, el matching por K vecinos más cercanos y por puntaje de propensión ponderado.

Gracias a la propiedad de balanceo, la selección a tratamiento es aleatoria al controlar por el puntaje de propensión. Por ello observaciones de control y tratamiento son en promedio idénticas, siempre y cuando, tengan el mismo puntaje de propensión.

Antes de terminar esta sección, podemos mencionar algunas recomendaciones. En primer lugar, controlar cual es el rango de puntajes de propensión para los grupos de control y tratamiento, de forma de utilizar un rango en el que se encuentren ambos grupos, descartando el rango en el que solamente se encuentra uno de ellos.

Generar histogramas con el puntaje de propensión estimado para los tratados y control y generar contenedores que correspondan con la estratificación a utilizar, si es que se utiliza esa técnica, para que idealmente la frecuencia de tratamiento y control sea similar en cada uno de los contenedores.

3.5.2 Matching mediante estrategia de estratificación

Esta técnica consiste en estratificar el puntaje de propensión en bloques de forma que dentro de cada bloque los grupos de tratamiento y control no sean estadísticamente diferentes. Una técnica sugerida es empezar con cinco bloques de igual rango y testear si la media de los puntajes es estadísticamente diferente. Si lo es, aumentar la cantidad de bloques y repetir. Una vez que se encuentra la cantidad que cumple con esta consigna, testear el balanceo de todas las variables observacionales X. Si se encuentra que alguna variable tiene diferencias estadísticamente significativas seguir dividiendo los bloques hasta lograr el balance.

Una vez cumplidas estas condiciones, calcular para cada uno de los bloques el ATE:

$$\tau_q^S = \frac{\sum_{i \in I(q)} Y_i^T}{N_q^T} - \frac{\sum_{j \in I(q)} Y_j^C}{N_q^C}$$

Donde $I(q)$ son las unidades que pertenecen al bloque q mientras que N_q^T y N_q^C son la cantidad de unidades de tratamiento y control en el bloque q .

El efecto promedio de tratamiento en los tratados (ATT) se calcula de la siguiente manera:

$$\tau^S = \sum_{q=1}^Q \tau_q^S \frac{\sum_{i \in I(q)} D_i}{\sum_{\forall i} D_i}$$

El peso que tiene cada bloque depende de la fracción de unidades que contiene cada uno.

3.5.3 Matching mediante estrategia de K vecinos más cercanos

En esta situación nos gustaría hacer matching de cada unidad de tratamiento con una de control que tenga exactamente el mismo puntaje de propensión. Como se ha mencionado anteriormente, esto es casi imposible. Lo más cercano a un matching exacto es emparejar a cada unidad de tratamiento con la unidad de control más *cercana* en términos del puntaje de propensión.

3.5.4 Estimación del ATE mediante la ponderación del puntaje de propensión

La utilización del puntaje de propensión nos permite prescindir del uso de las variables observacionales como controles ya que el mismo propensity score se transforma en el control. Se detalla la explicación matemática de esta afirmación:

Si la asignación al tratamiento no tiene valores de confusión, entonces $Y_{(1)}, Y_{(0)} \perp D | X$ y ATE es igual a:

$$ATE = E\{Y_i(1)\} - E\{Y_i(0)\} = E\left\{\frac{Y_i D_i}{p(X_i)}\right\} - E\left\{\frac{Y_i(1 - D_i)}{1 - p(X_i)}\right\}$$

Prueba matemática del enunciado anterior mediante la ley de probabilidad total:

$$E\left\{\frac{Y_i D_i}{p(X_i)}\right\} - E\left\{\frac{Y_i(1 - D_i)}{1 - p(X_i)}\right\} = E\left\{E\left\{\frac{Y_i D_i}{p(X_i)} \middle| X\right\} - E\left\{\frac{Y_i(1 - D_i)}{1 - p(X_i)} \middle| X\right\}\right\}$$

$$E \left\{ E \left\{ \frac{Y_i D_i}{p(X_i)} \mid D_i = 1, X \right\} Pr\{D_i = 1 \mid X\} - E \left\{ \frac{Y_i(1 - D_i)}{1 - p(X_i)} \mid D_i = 0, X \right\} Pr\{D_i = 1 \mid X\} \right\}$$

Utilizando la definición del puntaje de propensión y el hecho de que la ausencia de factores de confusión hace que el condicionamiento al tratamiento sea irrelevante en las dos esperanzas internas llegamos a la definición de ATE:

$$E\{E\{Y_i(1)|X\} - E\{Y_i(0)|X\}\} = E\{Y_i(1)\} - E\{Y_i(0)\} = ATE$$

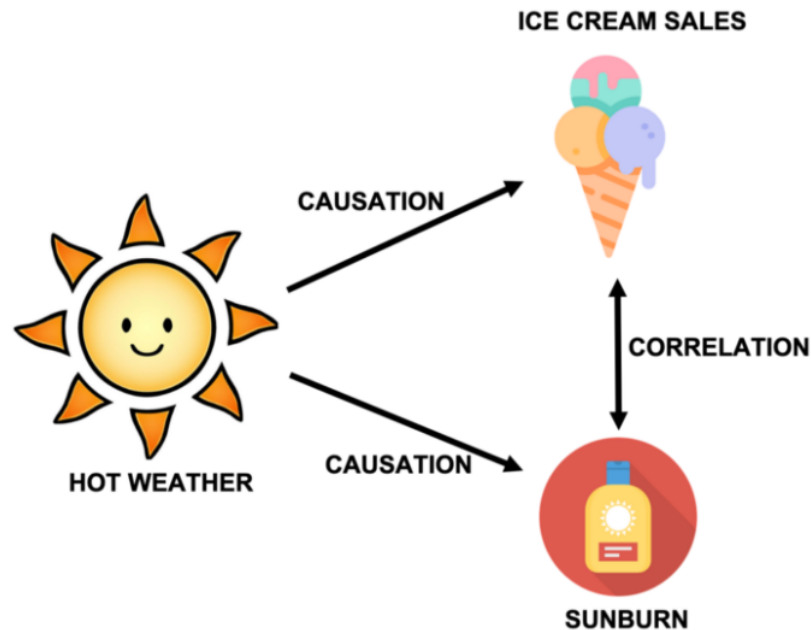
Esto significa que para las observaciones sin tratamiento ($D_i=0$) la esperanza equivaldrá a $\frac{Y_i}{1-p(X_i)}$. En el caso de las observaciones con tratamiento la esperanza equivaldrá a $\frac{Y_i}{p(X_i)}$. La ventaja del método de ponderación por el puntaje de propensión es que no es necesaria la estratificación ni los procedimientos de matching. Al hacer una prueba de impacto se recomienda aplicar más de un método y compararlos. Si se encontraran grandes diferencias en los resultados podría deberse a una mala estimación del puntaje de propensión o en una falla en la suposición de ausencia de factores de confusión.

3.6 Diferencia con técnicas de Machine Learning

Puede ser de interés del lector el entender las ventajas y diferencias entre realizar un estudio de inferencia causal y el de una predicción mediante métodos de aprendizaje automático, comúnmente conocidos como "Machine Learning". El aprendizaje automático es muy eficaz en responder preguntas de predicción, pero no todos los algoritmos de machine learning son adecuados para evaluar causalidad. Por otro lado, existen algoritmos que si pueden evaluar causalidad como las redes bayesianas, las cuales consisten en modelos causales gráficos y en la probabilidad condicional asociada a ellos. En este caso no solo se predice la asociación entre variables si no también el efecto que tiene una sobre otra.

La siguiente imagen es muy didáctica para transmitir la diferencia entre asociación y causalidad:

Imagen 4: Ejemplo causalidad vs correlación (Lovejoy, 2021)

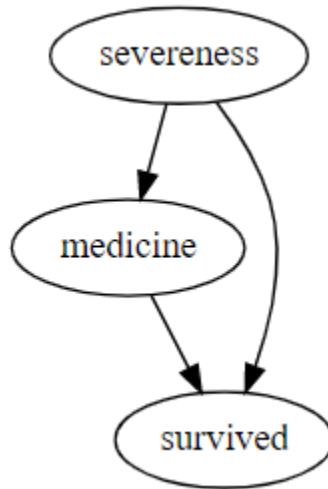


3.7 Modelos Causales Gráficos

Los Grafos acíclicos dirigidos, DAGs por sus siglas en inglés, son representaciones visuales de supuestos causales. Los datos por sí mismos solo comunican asociación, la cuál está compuesta por componentes causales y por componentes no causales. Los DAGs son una herramienta gráfica más accesible que la notación algebraica para comunicar relaciones de causalidad entre variables.

Se utilizará lo expresado por (Elwert, 2013) y (Alves, 2020) para explicar este concepto de manera concisa. Los DAGs consisten en tres elementos: variables (expresadas como nodos o vértices), flechas y la ausencia de flechas. La ausencia de flechas se destaca como elemento ya que es más importante que la presencia de estas. La razón es que las flechas expresan la *posible* relación causal entre pares de variables, mientras que la ausencia expresa el supuesto de que no existe relación causal entre las mismas.

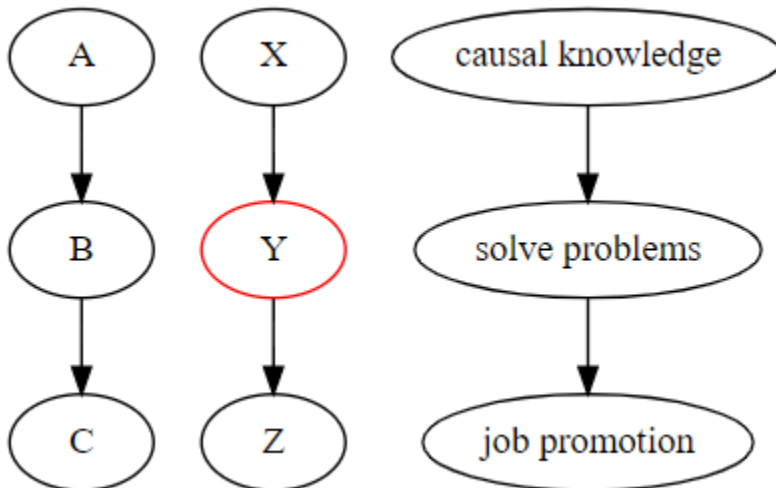
Imagen 5: Ejemplo de DAG (Alves, 2020)



Si bien existen infinitas combinaciones de variables y relaciones posibles, el foco será puesto en tres estructuras:

3.7.1 Cadenas

Imagen 6: Ejemplo de cadenas (Alves, 2020)



Las flechas indican el flujo de causalidad, por lo que A y B están asociadas por causalidad. C es una causa indirecta de A. En el ejemplo más concreto, saber sobre causalidad permite resolver problemas, y resolver problemas es la única manera de lograr un ascenso laboral. Por lo que el ascenso laboral es dependiente del conocimiento en causalidad. Al aumentar el conocimiento en causalidad, aumenta la chance de promoción.

Sin embargo, si condicionamos por la variable intermediaria, bloqueamos la dependencia. Por lo que X y Z serían independientes para un dado valor de Y. Siguiendo el ejemplo anterior, si soy bueno resolviendo

problemas, saber que tengo conocimiento en causalidad no me da más información sobre las chances de tener un ascenso laboral.

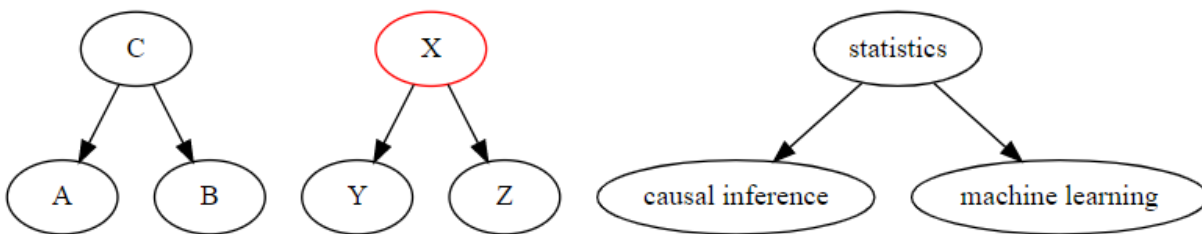
Matemáticamente se expresa cómo:

$$E[\text{Promoción} | \text{Resolver problemas}, \text{Conocimiento}] = E[\text{Promoción} | \text{Resolver Problemas}]$$

El flujo de dependencia puede ser bloqueado al condicionar por una variable intermedia.

3.7.2 Bifurcación

Imagen 7: Ejemplo de Bifurcación (Alves, 2020)

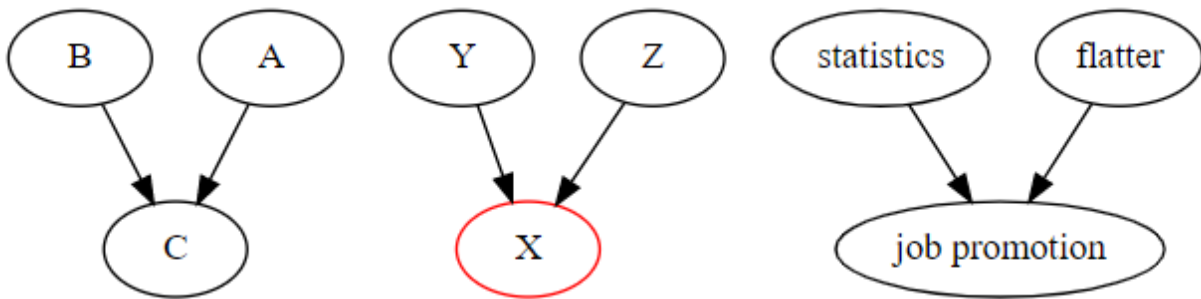


En este caso, una misma variable causa otras dos a lo largo del grafo. Al igual que en el caso anterior, si controlamos por la variable X que influye en las otras dos, podemos independizar la relación que hay entre Y y Z. El caso ejemplo es que el conocimiento en estadística influye en el conocimiento en inferencia causal y en el conocimiento en machine learning. Si no se controla el conocimiento en estadística, se puede llegar a inferir que alguien con conocimiento en inferencia causal también pueda tener conocimiento en machine learning. Pero, si controlamos por el conocimiento en estadística, lo que esa persona sepa de inferencia causal será independiente de su conocimiento en machine learning. Como regla, dos variables que dependen de una misma causa son dependientes entre ellas, a menos que controlemos por la causa común y así se vuelven independientes.

Esta asociación común entre dos variables es también conocida como “causa común de sesgo de confusión”.

3.7.3 Colisión

Imagen 8: Ejemplo de Colisión (Alves, 2020)



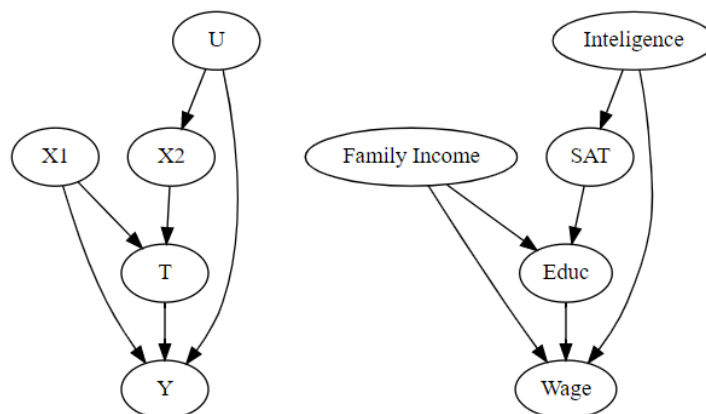
Esta estructura se genera cuando dos flechas coinciden sobre una misma variable. El ejemplo indica que hay dos maneras de obtener una promoción: Saber de estadística o halagar al jefe. En este caso saber el conocimiento de estadística de un caso no me da información acerca del nivel de halago que dirige al jefe. Sin embargo, si sé que tuvo un ascenso, y a la vez sé que no es bueno adulando al jefe, entonces puedo inferir que es bueno en estadística.

En este caso A y B son marginalmente independientes, ya que A no interfiere sobre B. Aunque, si controlamos por el resultado común X, Y y Z están asociadas por un resultado en común (el colisionador).

3.7.4 Sesgo de Confusión

Una de las grandes causas de sesgo es la “confusión”, conocida como “sesgo de confusión”. Este ocurre cuando el tratamiento y el resultado comparten una causa común. En este caso es difícil medir cuánto del efecto en el resultado depende del tratamiento y cuánto depende de la otra causa en común entre ambas. Un caso que puede transmitir la complejidad de este tipo de análisis es el efecto de la educación en los salarios:

Imagen 9: Ejemplo de Sesgo de Confusión (Alves, 2020)



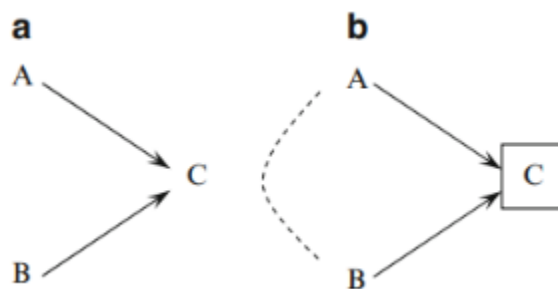
El salario en este ejemplo depende directamente del ingreso familiar, de la educación y también de la inteligencia, la cual adiciona otra complejidad que es que no es fácil de medir. Sin embargo, sí podemos controlar por SAT, que es el resultado en exámenes estandarizados para el ingreso a universidades. Si controlamos por X1 y X2, es decir, SAT e Ingreso familiar, podemos obtener independencia condicional de la variable que no podemos medir (Inteligencia) sobre la variable tratamiento (Educación) y así eliminar el sesgo que aporta la inteligencia como variable de confusión.

En los ejemplos reales no siempre es posible eliminarla completamente, ya que la variable de confusión podría también inferir directamente sobre la variable tratamiento, por lo que en esos casos al controlar por las otras variables que afectan al tratamiento lograríamos disminuir el sesgo de confusión, pero no eliminarlo totalmente.

3.7.5 Sesgo de Selección

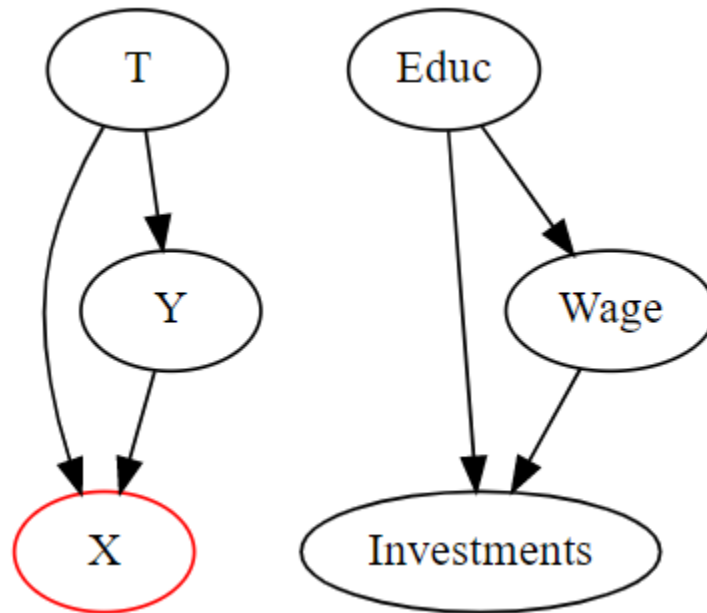
Otro tipo de sesgo posible es el de selección. Al contrario que en el sesgo de confusión, que ocurre por no controlar por variables que impactan sobre tratamiento y resultado, en el caso del sesgo de selección ocurre por controlar por más variables de las que deberíamos. En este caso, si controlamos por variables que son causadas por el tratamiento y el resultado, entonces lo que hacemos es crear una relación entre tratamiento y resultado que va a hacer más difícil distinguir el efecto directo.

Imagen 10: Sesgo de Selección (Elwert, 2013)



En el caso **a**, las variables A y B son independientes, pero si condicionamos por la variable consecuencia de ambas C, generamos una relación entre A y B lo cual nos lleva al escenario de sesgo de confusión entre ambas. Esto se puede observar con mayor claridad en el siguiente ejemplo:

Imagen 11: Ejemplo de sesgo de selección (Alves, 2020)



Queremos evaluar el impacto que tiene la educación en los salarios. Al ser consecuencia de ambas, si controlamos por inversiones lo que hacemos es indirectamente controlar los salarios, ya que los salarios son causa de las inversiones. Al crear pequeños grupos donde los salarios están en mayor medida fijos por ser dependientes del nivel de inversión, no le estamos permitiendo cambiar lo suficiente y no será posible observar el efecto que tiene la educación para cambiar el salario.

Es decir, quienes invierten pero no tienen educación, necesariamente tienen que tener altos salarios. Sin embargo, quienes tengan alta educación pueden llegar a tener bajos salarios. De esta manera, por sesgo de selección, estamos asociando alta educación con bajos salarios.

Lo desarrollado en esta sección teórica es útil para comprender la diferencia entre asociación y causalidad. Al mismo tiempo se explicaron brevemente las diferentes estructuras que pueden encontrarse en modelos causales gráficos y los sesgos que pueden generarse si no se controlan las variables, o si por el contrario se controla de más e incorrectamente.

4. Datos y Materiales

En este trabajo utilizaremos un set de datos público⁶ proporcionado por el Ministerio de Educación de la República Argentina en el cual se exponen los resultados de las pruebas Aprender del año 2018. Los datos fueron descargados en el año 2020 del mismo sitio. De acuerdo con el Ministerio de Educación, ‘Aprender’ “es el dispositivo nacional de evaluación de los aprendizajes de los estudiantes y de sistematización de información acerca de algunas condiciones en las que ellos se desarrollan.”

Tanto el código como los datos utilizados como así fueron descargados en 2020 para llevar a cabo este estudio se encuentran en la carpeta pública del citado Google drive.⁷

Los datos comprenden los resultados de un cuestionario hecho a los alumnos de sexto grado en 2018 junto con los resultados de un examen de lengua y matemática. Por lo que al mismo tiempo brinda información sobre el alcance de conocimientos y competencias, y también información de contexto de los alumnos. Los datos no son una muestra, sino que comprenden al universo completo de estudiantes de sexto grado.

En total, tenemos las respuestas y resultados de 579.510 alumnos en más de 100 preguntas que van desde descriptores básicos como la edad y el género, a otras relacionadas con la cantidad de personas con las que conviven, el nivel educativo de sus padres, el acceso a internet, a celulares o computadoras. Se indaga sobre las diferentes tareas hogareñas o laborales que realiza el alumno, su pasado escolar, su percepción de la escuela, su autopercepción en distintas aptitudes académicas y sobre el índice socioeconómico del alumno. En cuanto a la escuela se detalla si pertenece a la gestión estatal o privada, el ámbito urbano o rural, la provincia a la que pertenece y contexto social en el que se da la educación (Bajo, Medio, Alto).

Esta información será utilizada para el análisis de inferencia causal, a fin de determinar si el acceso a una computadora impacta positivamente en los resultados escolares, separando el efecto de otras variables de confusión.

⁶ <https://www.argentina.gob.ar/educacion/aprender2018>

⁷ <https://drive.google.com/drive/u/1/folders/1EapWHWfJOWzcnTm22xew03QEo6BZWXW>

5. Análisis Descriptivo

El set de datos cuenta con información muy completa respecto a la población de alumnos de sexto grado de la República Argentina. Se analizarán las diferentes variables con el objetivo de extraer información relevante previo al estudio de causalidad.

5.1 Tratamiento Missings

Los datos contienen registros de 579.510 alumnos y es de esperar que exista información faltante en algunas de ellas, por lo que es necesario establecer un marco de referencia para el tratamiento de faltantes en los datos. En particular, en los resultados de Lengua y Matemática se observan frecuencias elevadas de missing value del 3,92% y 4,74% respectivamente. En las variables que serán controles se observan algunos missings en variables categóricas que serán considerados como una categoría diferente. En todos los casos se mantiene dentro del ejercicio todas las observaciones sin excluir ninguna por no estar completo su registro. En el estudio de impacto es necesario definir qué variables serán controles, cuál será tratamiento y cuál respuesta (outcome). El acceso a una computadora en el hogar será nuestra variable de tratamiento, ya que es sobre la cual queremos inferir el efecto, y los resultados en lengua y matemática serán nuestras variables de respuesta.

El marco para el tratamiento de los fs tiene como base a (Li, 2013). En el referido estudio, se analizan los posibles mecanismos por el cual existe falta de información en los sets de datos. Siendo R una variable Y (respuesta) que denota la falta de información en una variable de forma que $R=0$ signifique que falta Y ; y $R=1$ que se observa Y . En nuestro caso, Y representa el resultado del examen de lengua o matemática según corresponda.

Existen entonces 3 escenarios respecto a los missings en nuestra variable de outcome (Y).

- **MCAR:** Missing Completely at Random. Significa que la falta de información es debido al azar y completamente independiente de cualquier variable. Expresado matemáticamente: $R \perp X, Y$
- **MAR:** Missing at Random. Bajo MAR, la falta de información depende únicamente de variables observables. De esta manera $P(R=0|X, Y) = P(R=0|X)$

- **MNAR:** Missing Not at Random. Cuando no se encuentra bajo MCAR o MAR, estamos bajo MNAR en el que hay variables no observables asociadas a la existencia del missing. Es posible que variables observables también tengan influencia. Un caso posible es en las encuestas referidas a fenómenos sociales en el que el encuestado decide activamente omitir responder a algunas preguntas por sus propias razones. En estos casos la falta de respuesta es una respuesta en sí misma, y eliminar todas las respuestas del encuestado por esta razón implicaría una pérdida de información.

Para contemplar un posible sesgo de selección se realizará una corrección por los missings en la estimación del ATE. En este caso, bajo el supuesto de MAR, consideraremos los missings como parte del mecanismo de selección muestral. Siguiendo la tradición del inverse probability weighting method (IPW) propuesto por (Horvitz & Thompson, 1952) ponderamos por la probabilidad de tener completo el puntaje del examen condicional a los controles (X) bajo el supuesto del cumplimiento del supuesto de MAR. (Robins, Rotnitzky, & Zhao, 1995) prueban que el estimador de inverse probability weighting es consistente y asintóticamente normal en un modelo de regresión multivariada y proponen un estimador IPW modificado (Little, 1995). (Han, 2018) provee un intervalo de confianza asintótico para este estimador del ATE. Entonces, el estimador del ATE que corrige por el sesgo de selección potencialmente generado por los missings en la variable outcome resulta:

$$\hat{\delta} = N^{-1} \sum_{i=1}^N \left[\frac{R_i W_i Y_i}{p(X_i) p(R_i = 1|X_i)} - \frac{R_i (1 - W_i) Y_i}{(1 - p(X_i)) p(R_i = 1|X_i)} \right]$$

5.2 Datos generales de los alumnos

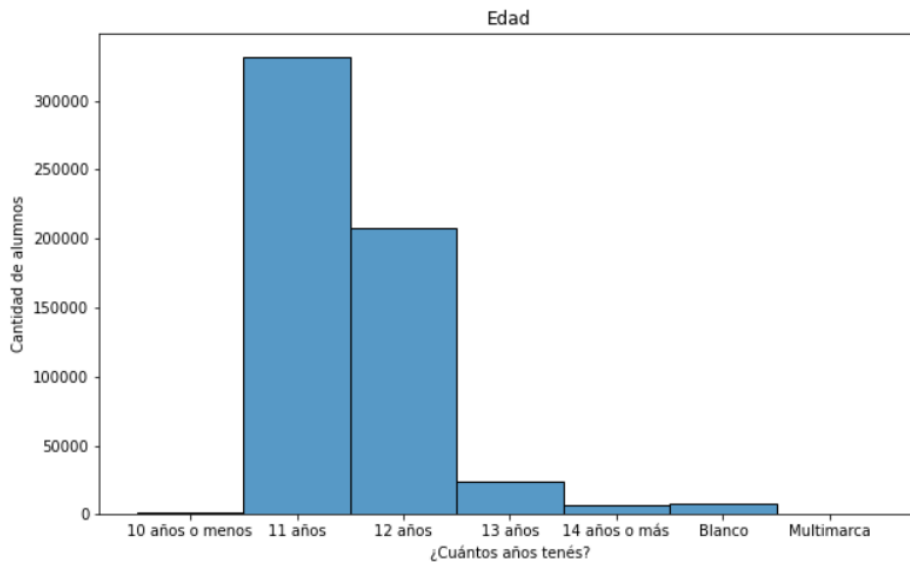
El 96% de los alumnos nacieron en el país, siendo los dos países extranjeros con mayor volumen Paraguay y Bolivia. El dato es de interés ya que en los exámenes PISA 2015 detectaron que los alumnos con antecedente inmigrante, definidos como alumnos cuyos padres nacieron en otro país del que actualmente habitan, tienen el doble de propensión a tener un desempeño por debajo del nivel 2 en ciencia.⁸ Existe una diferenciación entre alumnos inmigrantes de primera generación que son quienes tanto como sus

⁸ OECD, "Immigrant background, student performance and students' attitudes towards science", 2016

padres son inmigrantes, y los de segunda generación en los que ambos padres son inmigrantes pero el alumno es nacido en el país que habita.

La distribución por Género es muy equitativa. El género es un dato para tener en cuenta ya que en los exámenes PISA de 2018 se observó que la distribución de género en los resultados promedio de los exámenes de lengua no era significativa, pero en los extremos las diferencias de género se acentuaban. (OECD, "Girls' and boys' performance in PISA", 2020)

Figura 2: Cantidad de alumnos por edad



La distribución por edad se encuentra dominada por la franja de 11 a 12 años que es la esperada en niños de sexto grado. Sin embargo, se detecta una porción de la población que tiene 13 años y otra que tiene 14 años o más. Esto plantea dos posibilidades: que esos alumnos hayan repetido el año previamente, o que hayan ingresado al colegio más tarde de lo esperado. Ambos casos son para considerarlos como de mayor riesgo ya que el mal rendimiento escolar previo es compatible con el modelo de autoestima planteado por Finn, y podría derivar en un círculo vicioso de mal rendimiento escolar. (Finn J. D., 1989)

5.3 Datos Socioeconómicos

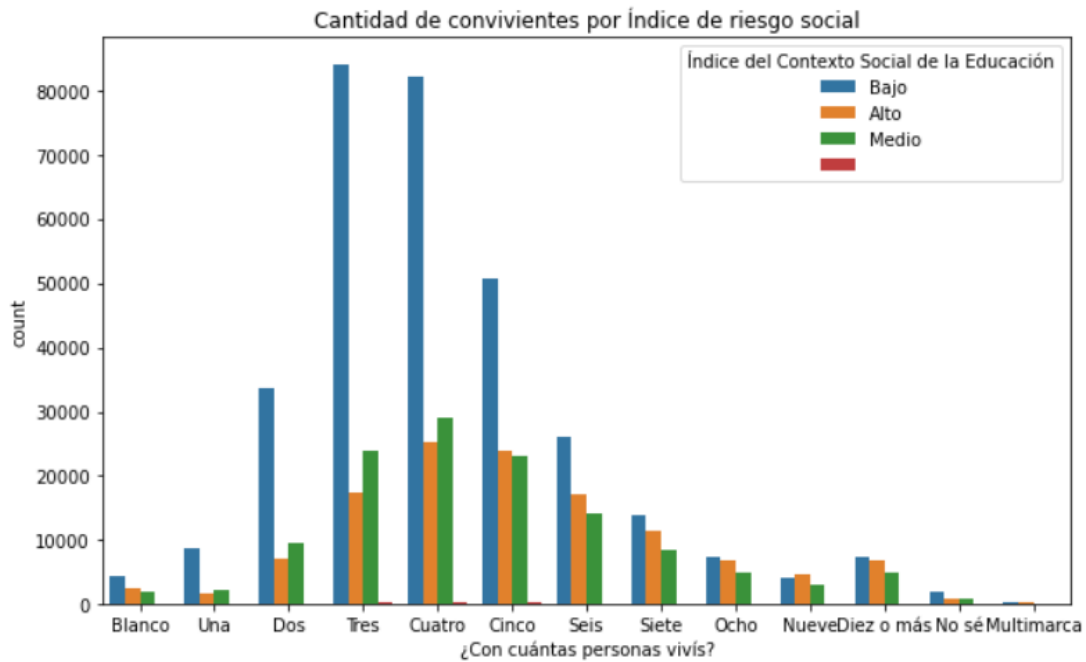
La información más valiosa del set de datos está conformada por las preguntas referidas a cómo vive cada uno de los alumnos de Argentina. Esta información va desde cómo está conformado el núcleo familiar, al nivel educativo de sus padres y a las comodidades o posibilidades materiales a las que están acostumbrados.

La cantidad de personas que conviven en el hogar con el alumno encuentra su mayoría de respuestas en tres a cuatro personas, y luego va decreciendo proporcionalmente a medida que aumenta la cantidad de personas. De esta dimensión de datos se pueden hacer muchos análisis e hipótesis. En cuanto a hipótesis podemos plantear que la respuesta de 3 a 4 convivientes podría ser el núcleo familiar tradicional conformado por padres y hermanos. A medida que la cantidad de convivientes aumenta, esto puede significar que la cantidad de hermanos está por encima de la media, o que lo que ocurre es que hay más miembros de la familia viviendo juntos, como pueden ser tíos o abuelos.

Al mismo tiempo este desglose genera la pregunta de si efectivamente las familias más numerosas son las de menores recursos, mientras que las más ricas o de bajo riesgo socioeconómico se encuentran entre las menos numerosas. Ambas preguntas serán respondidas a lo largo de la exploración de datos.

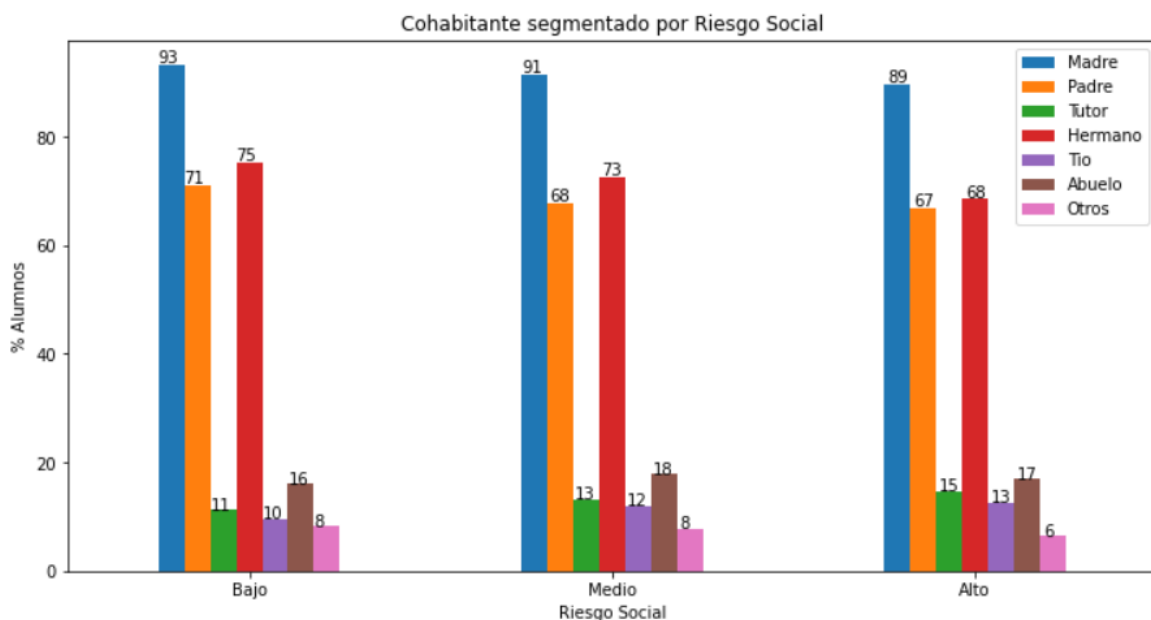
Por suerte podemos responder estas preguntas gracias al 'Índice del Contexto Social de la Educación' que está proporcionado dentro del mismo set de datos. El mismo está definido por el Ministerio de Educación de la siguiente manera: *'El Índice del Contexto Social de la Educación fue desarrollado por el Ministerio de Educación, Cultura, Ciencia y Tecnología. El mismo utiliza la información censal para la clasificación de pequeñas unidades geográficas en base a una aproximación multidimensional a las condiciones de vida y situación de vulnerabilidad de los hogares con residentes en edad escolar. Este índice permite describir el nivel de vulnerabilidad del radio censal en el cual se ubican las escuelas.'* (Ministerio de Educación, Cultura, Ciencia y Tecnología de la República Argentina, 2019)

Figura 3: Cantidad de alumnos por cantidad de personas convivientes según índice del contexto social de la educación



Del gráfico podemos inferir cómo aumenta la proporción de alumnos de vulnerabilidad alta y media en las familias más numerosas. La pregunta de quiénes son los miembros del núcleo familiar se puede responder con la siguiente información:

Figura 4: Cantidad de alumnos por familiar conviviente según índice de contexto social de la educación

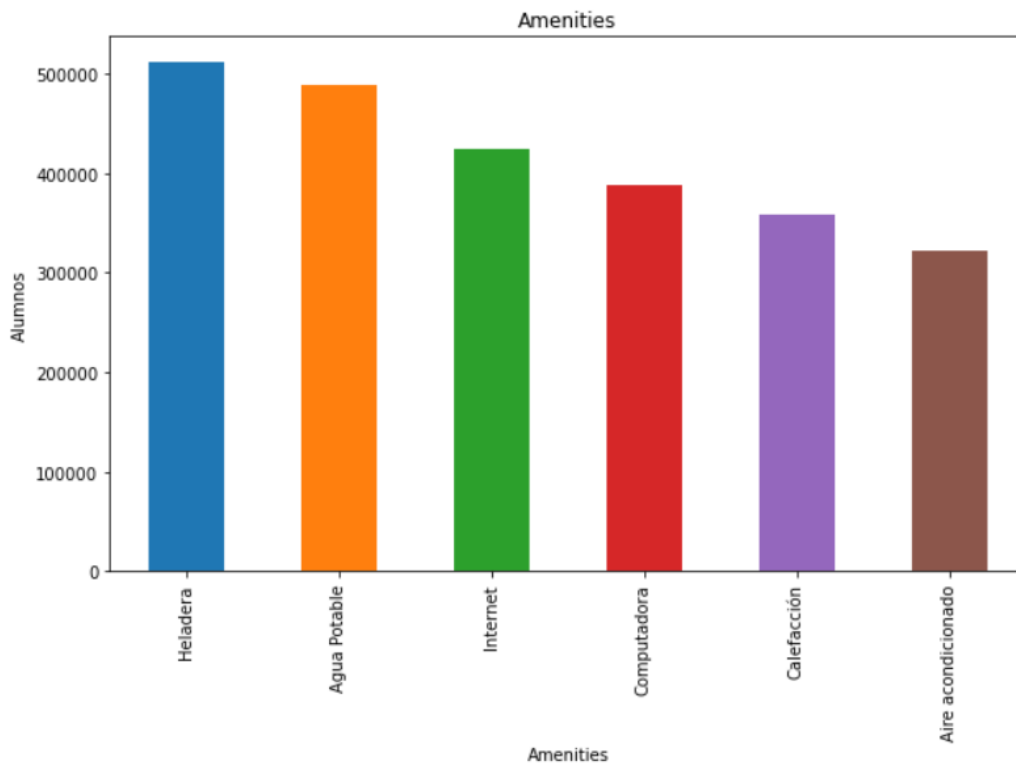


Casi todos los alumnos viven con su madre, y luego una gran mayoría vive con sus hermanos y padre. Resta luego una minoría que vive junto con sus abuelos, tutores, tíos y otros. Explorando este dato y diferenciando por el índice de contexto social se puede observar que la diferencia entre ellos no es tan marcada. Es cierto que existe un declive en el porcentaje de convivientes con padre, madre y hermanos, los cuales componen la familia tradicional, y también hay un incremento en el porcentaje de alumnos que conviven con tíos y tutores. Sin embargo, no se ve alterada drásticamente como está conformada la familia. De acuerdo con la información analizada hasta ahora, podemos concluir que los alumnos de menores recursos conviven en promedio con una persona más en su casa que los de riesgo social bajo. Esta persona extra tiene mayores chances de ser un tutor o un tío, en comparación a los otros segmentos.

5.3 Condiciones habitacionales

Debajo se visualizarán y analizarán las condiciones habitacionales de los alumnos. Es de esperar que el grupo de mayor riesgo cuente con menor cantidad de comodidades. Estas variables actuarán como controles, excepto la computadora que será la variable tratamiento.

Figura 5: Cantidad de alumnos por comodidad del hogar



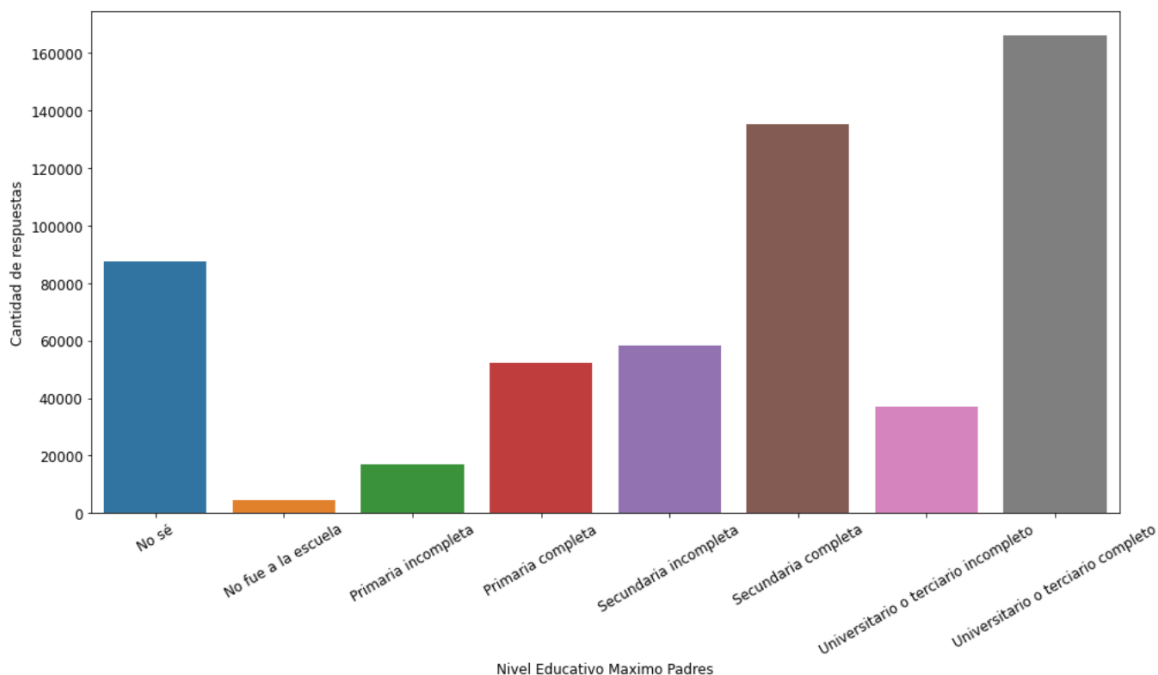
Los elementos de climatización parecieran ser los que más escasean, mientras que Heladera y Agua Potable son los que más abundan dada su condición de ser de básica necesidad. Es en el centro de la distribución que se encuentra Internet y Computadora. Dos elementos estrechamente relacionados, y que son herramientas muy utilizadas para el estudio y el ocio. Se dejará planteada aquí una de las preguntas más relevantes de este estudio: **¿Qué impacto tiene la posesión de una computadora en las notas finales de los alumnos?**

5.4 Nivel educativo y cultural familiar

Otro factor ambiental relevante para el alumno es el nivel educativo de sus padres. Esto implica tanto el nivel de educación formal al que hayan llegado, como así también la cantidad de libros que haya en la casa.

La pregunta acerca de la cantidad de libros que hay en el hogar tiene 23% de respuestas 'no sabe'. En el caso de los chicos que sí han especificado una cantidad, se destaca que un 16% de los alumnos indican que afirmativamente no hay libros en la casa. El valor máximo se da en la respuesta de 1 a 25 libros con el 24% de las respuestas.

Figura 6: Cantidad de alumnos según máximo nivel educativo de los padres



El nivel educativo de los padres originalmente viene desglosado por padre y madre. Para obtener la imagen más precisa de la educación disponible en el hogar, siguiendo la literatura (Davis-Kean, 2005), se utilizará el nivel más alto de los padres como indicador.

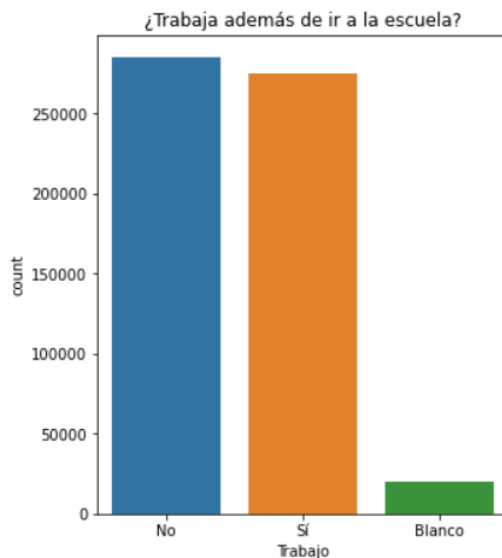
Los resultados indican que la mayoría de los padres cuentan con el nivel universitario o terciario completo. Seguido por los que tienen la secundaria completa. El porcentaje de padres que no completaron el nivel secundario ronda el 24%. Se destaca que un 15% de los alumnos desconoce el nivel educativo de sus padres.

La cantidad de libros en la casa como el nivel educativo de los padres son variables relevantes ya que existe una teoría de “movilidad cultural” que ve el capital cultural como proceso activo en el que las personas lo adquieren mediante interacciones, especialmente entre padres e hijos en el hogar. Esta postura sugiere que el ambiente social del hogar es un punto de partida (Carolán & Wasserman, 2014).

5.5 Trabajo y tareas hogareñas

El hecho de trabajar implica que el alumno tiene otras obligaciones y responsabilidades aparte de las escolares, las cuales pueden llegar a estar en conflicto con estas. El foco será puesto en sí los chicos trabajan o no, por lo que en esta sección se responderá la siguiente pregunta: ‘¿El alumno, además de ir al colegio, trabaja?’

Figura 7: Cantidad de alumnos que trabajan



Al combinar las respuestas de “trabajo con un familiar” y “trabajo fuera de casa” tenemos un resultado sorprendente. Casi la mitad de los encuestados respondieron que realizan algún tipo de trabajo además de ir a la escuela.

Este grupo podría considerarse de mayor riesgo y será tenido en cuenta en los análisis posteriores.

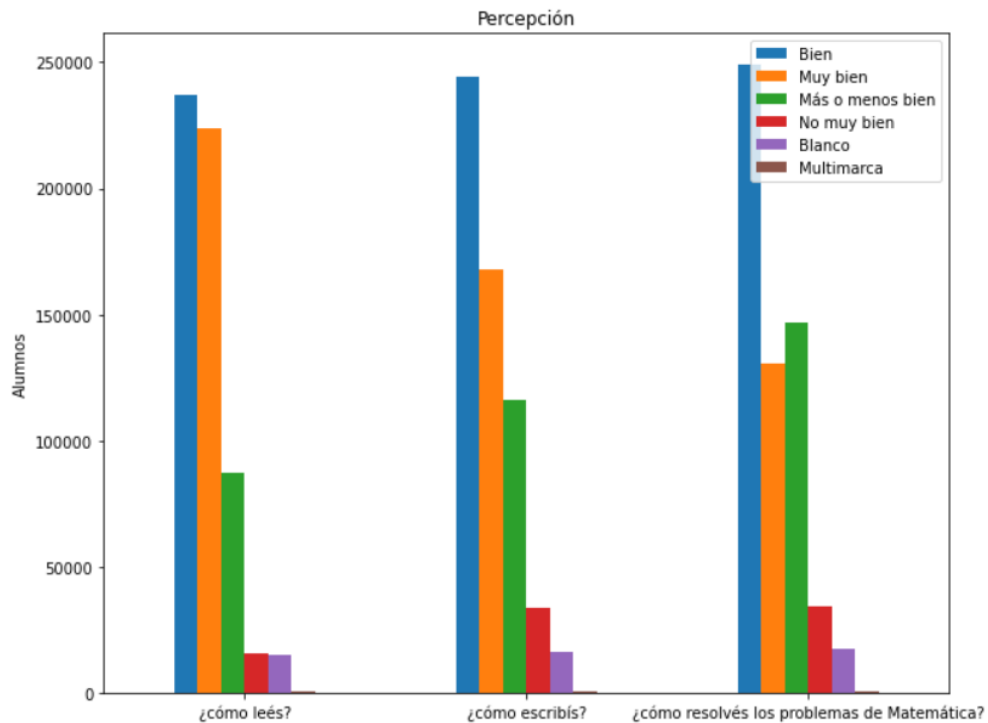
5.6 Compromiso y conexión con la escuela

La conexión y compromiso con la escuela son claves para evitar el abandono escolar de acuerdo con lo escrito por (Berrueta-Clement, 1984) y por (Finn J. D., 1989). En esta sección del análisis descriptivo revisaremos varias de las métricas relevantes a estos temas, como lo son la asistencia al jardín de infantes, la repetición del grado, el agrado de ir al colegio y la autopercepción del alumno para la lectura, escritura y matemática.

La asistencia al jardín de infantes es bastante dispar. La mitad asistió desde antes de los cuatro años, y el resto se reparte entre quienes fueron a partir de los cuatro y cinco años. Se destaca la existencia de un grupo pequeño de riesgo compuesto por el 1,4% de los alumnos que no fueron al jardín. Berrueta-Clement en la citada literatura describe cómo quienes asistieron al jardín de infantes mostraron mejores resultados académicos durante la primaria y la escuela media, tuvieron menores ausencias y a la vez mostraron tener una mayor valoración y compromiso por la escuela. Los efectos positivos se mantuvieron aún hasta el final de la secundaria.

Respecto a las repeticiones de año, el 87% no repitió nunca de grado. Los grupos que si repitieron serán considerados de riesgo ya que podemos considerarlos como de menor compromiso con la escuela, más aún teniendo en cuenta que hablamos de alumnos de primaria.

Figura 8: Auto percepción del rendimiento escolar

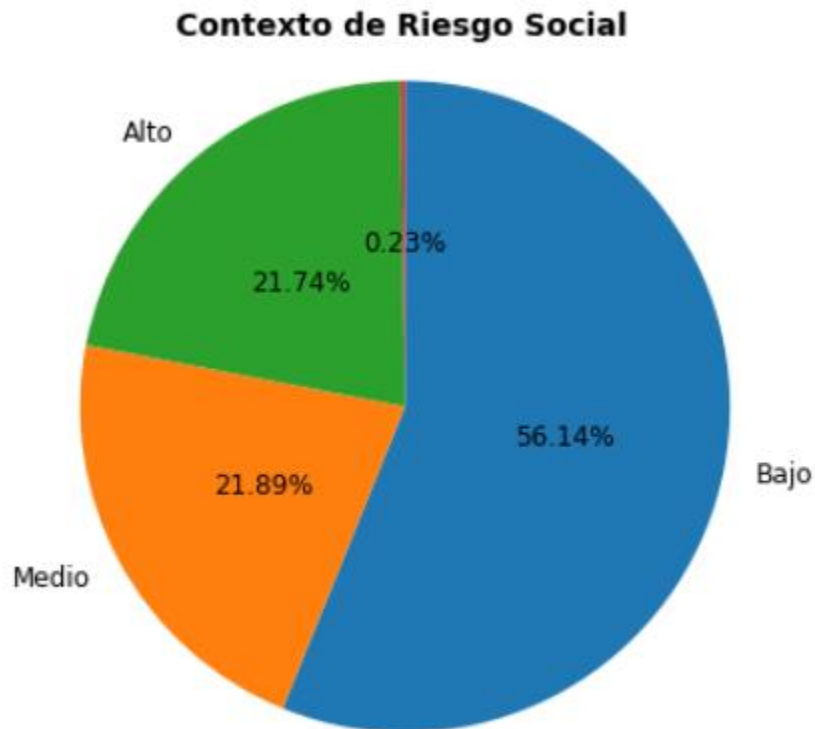


En la auto percepción del rendimiento escolar se puede observar que los alumnos son menos confiados en matemática, y a la vez que existe un grupo pequeño que reconoce que tiene problemas para leer, escribir o resolver problemas matemáticos. Barbara M. Byrne escribió sobre la auto percepción del alumno y el impacto que tiene efectivamente en sus calificaciones. En su investigación menciona que se han hecho varios estudios al respecto y que en ellos se han encontrado correlaciones positivas entre la auto percepción y el rendimiento escolar. A su vez, cita varios estudios en los que se ha estudiado la relación entre auto percepción y resultados académicos y los resultados son muy dispares. Una parte de la literatura sostiene que la percepción se genera a partir de la experiencia académica (resultados) y la otra que la auto percepción tiene una influencia en los resultados. (Byrne, 1984)

5.7 Ámbito de la Educación

Aproximadamente el 90% de los alumnos se educa en un ámbito urbano. El resto lo hace en ámbitos rurales. La educación en Argentina es en un 70% estatal, la cual es pública y gratuita y el resto es en establecimientos privados, los cuales pueden o no estar subsidiados por el estado y requieren el pago de cuotas mensuales para concurrir.

Figura 9: Distribución de alumnos según el contexto de riesgo social



Como se mencionó anteriormente, el riesgo social es una métrica que se calcula en torno a la geografía de la escuela. La información que obtenemos es que el 56% de los chicos tienen un riesgo social bajo, pero el restante 44% se divide en partes iguales entre un riesgo medio y alto. Esta clasificación será relevante a la hora de evaluar los resultados en las pruebas de matemática y lengua.

5.8 Resultados en Exámenes

El puntaje en los exámenes está expresado en valores que van del 241 al 760, clasificando la competencia de los alumnos en cuatro grupos distintos: “Por debajo del nivel básico”, “Básico”, “Satisfactorio” y “Avanzado”. La escala de puntajes es arbitraria y está pensada para que tenga media en 500 y un desvío estándar de 100.⁹

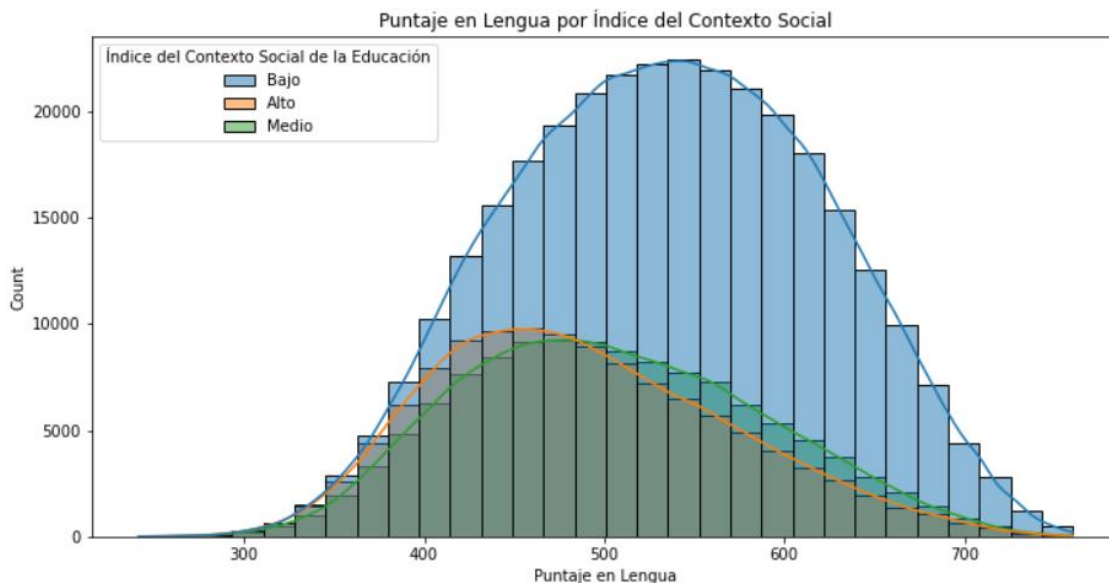
Tabla 1: Escala de desempeño según el puntaje en los exámenes

Nivel de Desempeño	Puntaje Mínimo	Puntaje Máximo
Por debajo del nivel básico	241	389
Básico	389	447
Satisfactorio	447	548
Avanzado	548	760

A continuación, se observa cómo cambia la distribución de calificaciones en el examen de lengua según el riesgo social.

⁹ El ministerio expresa en un documento técnico: “*En Aprender se escalaron los valores ϑ mediante una transformación lineal (multiplicando por un factor y sumando una constante) para que tenga media 500 y desvío standard 100, tomando Aprender 2016 como punto de referencia.*” (Ministerio de Educación, 2018)

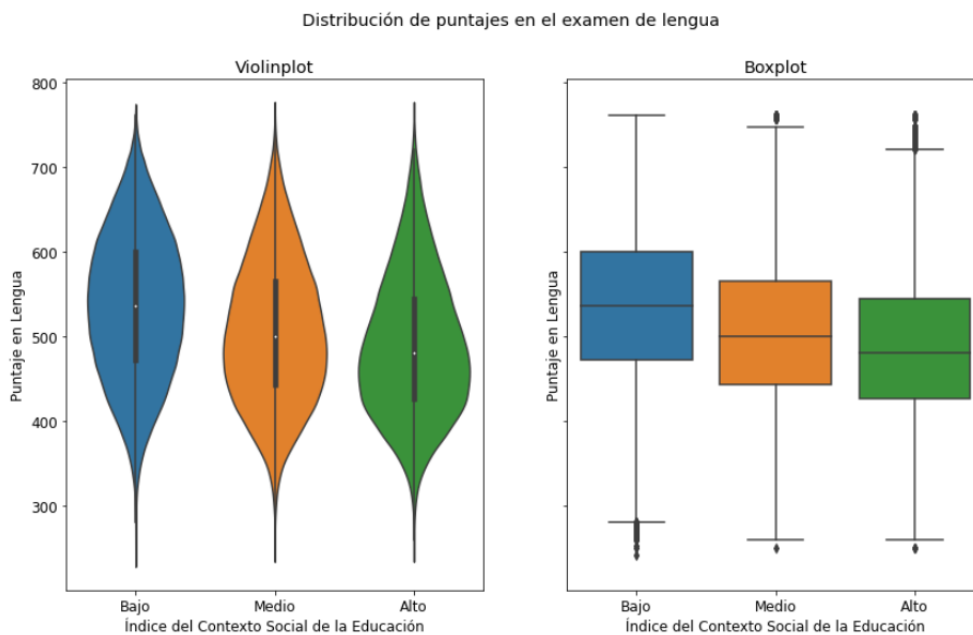
Figura 10: Distribución de puntajes en lengua según el índice del contexto social de la educación



Los resultados coinciden perfectamente con lo expuesto por Jeremy Finn (Finn J. D., 1989), en que desproporcionadamente el número de chicos que abandonan la escuela pertenecen a hogares de bajo nivel socio económico. En el gráfico podemos observar claramente como las calificaciones son mucho menores para los grupos de riesgo social medio y alto.

5.8.1 Resultados Lengua

Figura 11: Distribución de puntajes en lengua según el índice del contexto social de la educación



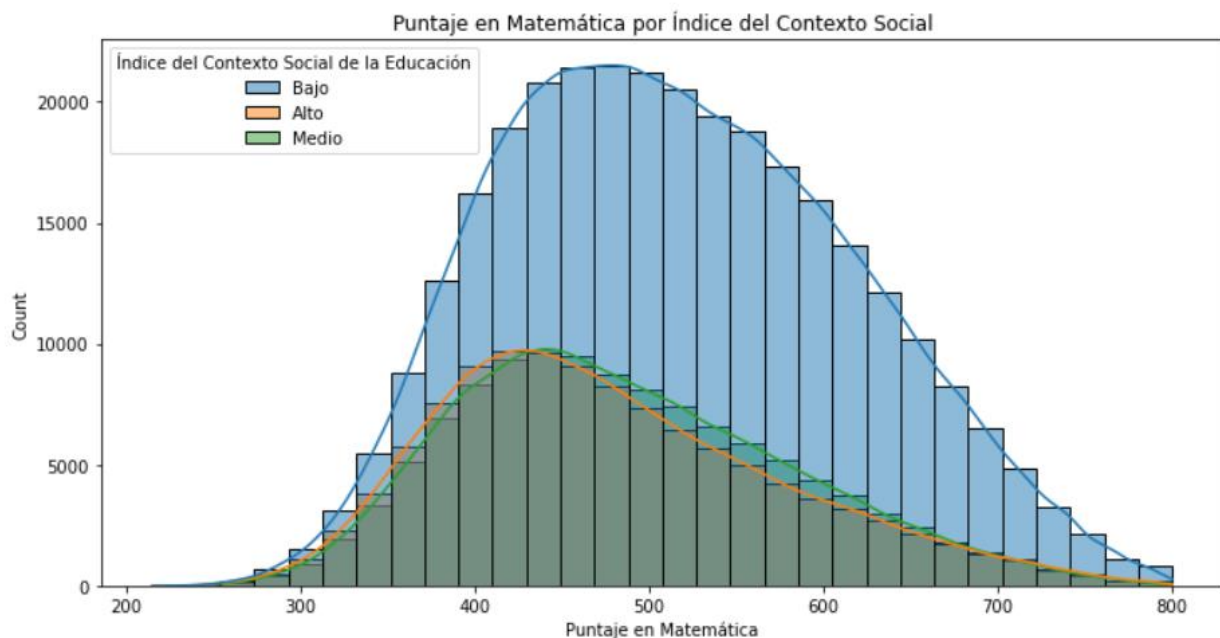
Se puede observar que a mayor riesgo social menor es el desempeño académico. El hecho de pertenecer a una clase social más vulnerable implica que las posibilidades académicas son menores ya que los que pertenecen a las categorías de riesgo social más alto tienen en conjunto rendimientos menores que los que están en riesgo social menor. Esto atenta también contra la posibilidad de ascender socialmente mediante la educación, ya que en estos resultados vemos que los mejores resultados los tienen quienes socioeconómicamente están mejor.

Lo cual lleva de nuevo a la pregunta inicial de este estudio. ¿Tener acceso a una computadora en el hogar impacta en el rendimiento escolar? De poder identificar el efecto que tiene tener una computadora como tratamiento se podría considerar intervenir para disminuir o evitar el círculo vicioso de que el mayor riesgo social sea retroalimentado con peores rendimientos académicos, lo cual reduce la posibilidad de ascenso social.

5.8.2 Resultados Matemáticas

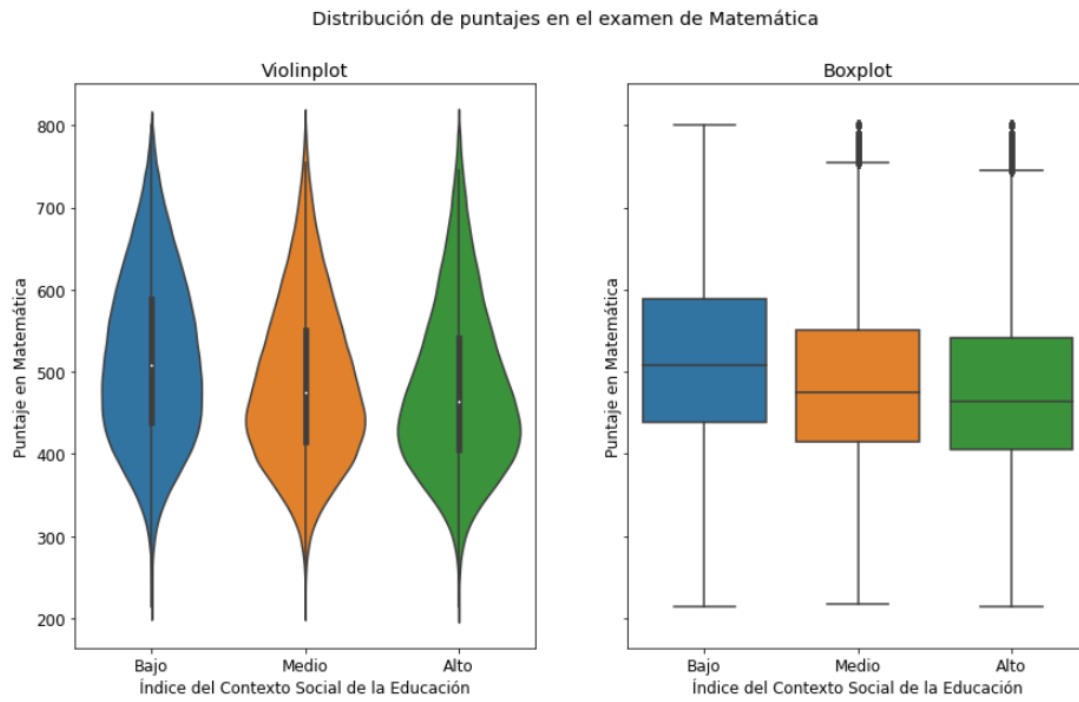
En los resultados en matemática también se observa la diferencia en la distribución de calificaciones entre los que tienen riesgo social bajo y los que tienen medio y alto.

Figura 12: Distribución de puntajes en matemática según el índice del contexto social de la educación



Los resultados son similares a los de lengua, pero con puntajes ligeramente menores:

Figura 13: Distribución de puntajes en matemática según el índice del contexto social de la educación



6. Estimación del Average Treatment Effect (ATE)

Como primer paso, se trabajó sobre los datos para normalizar las variables continuas como son los resultados en los exámenes, y se transformaron algunas variables para lograr una respuesta más estandarizada, como en la educación de los padres o si el alumno trabaja o no. Las variables que mencionan el máximo nivel educativo de padre y madre fueron agrupadas en “máximo nivel educativo de los padres”, las preguntas relacionadas a trabajo con la familia y fuera de la familia fue unificada en si trabaja o no. Las variables que refieren al país de origen del padre y la madre se transformaron en dos preguntas de verdadero o falso: “Ambos Padres Argentinos” y “Ambos Padres Extranjeros”.

No fueron eliminadas las respuestas vacías ya que es posible que el hecho de no responder una pregunta no sea aleatorio y se deba a algún motivo consciente de quien responde, por lo que todas las observaciones fueron conservadas.

Para el procesamiento de los datos fue necesario realizar un procedimiento de “one hot encoding” para transformar todas las preguntas con respuestas categóricas en columnas de verdadero o falso mientras que las variables continuas, en nuestro caso los resultados, se normalizan para poder obtener resultados expresados en desvíos estándar. Esto resultó en una tabla de 540.688 filas y 304 columnas

Se utilizó más de un método de análisis para así observar si el efecto de tratamiento promedio para cada uno de ellos varía según la técnica utilizada.

La variable tratamiento es el acceso a una computadora en el hogar: Se busca determinar el impacto que pueda tener en los resultados en lengua y matemática la posesión de una computadora en el hogar. Como se mencionó anteriormente, se busca aislar el efecto de esta variable de todas las demás.

6.1 Selección de variables

Como menciona (Gertler, Martínez, Premand, Rawlings, & Vermeersch, 2017), la estimación de la técnica de propensity score matching para evaluar efectos causales es tan buena como las variables utilizadas para su cálculo. Esto propone un grado de dificultad adicional para el investigador al tener que decidir qué variables incluye y cuáles descarta.

Es por ello por lo que además de tener dominio en el uso de técnicas de inferencia causal, también es requerimiento tener conocimiento en la materia en que se aplicará la técnica. Sin embargo, aun conociendo la disciplina en cuestión, es necesario comprender qué significa que una variable sea de ayuda

en la predicción. Alan Brookhart y otros investigadores, (Brookhart, et al., 2006), realizaron un estudio enfocado en la selección de variables para la construcción de modelos basados en puntaje de propensión. Destacan que, aún con la creciente popularidad de este tipo de modelos, existe escasa literatura al respecto de cómo seleccionar correctamente las variables que se utilizarán para su cálculo. Los autores mencionan que hay incertidumbre respecto a si el foco tiene que estar puesto en variables que estén relacionadas a la intervención, o si, por el contrario, es conveniente focalizar en variables que expliquen el resultado. Para resolver este dilema utilizan simulaciones de Montecarlo, a partir de las cuales aplicaron propensity score matching con diferentes conjuntos de variables para evaluar la calidad de la predicción del efecto causal. En este contexto, los autores evaluaron variables relacionadas a la intervención, variables relacionadas al resultado y variables que estén relacionadas a ambas. El estudio reveló que, si una variable está relacionada con la intervención, pero no tiene relación con el resultado, agregarla solo incrementa la varianza del resultado sin reducir el sesgo. Por lo que únicamente añaden “ruido” al estudio. Los resultados obtenidos apoyan la sugerencia de adherir a lo que teóricamente expusieron Rubin y Thomas (Rubin & Thomas, Matching Using Estimated Propensity Scores: Relating Theory to Practice, 1996)), que es que deben ser incluidas las variables que estén relacionadas con el resultado, inclusive si no tienen relación con el tratamiento.

6.2 Técnicas utilizadas

En esta evaluación buscamos estimar el ATE para el tratamiento “contar con computadora en el hogar”. Las técnicas que se utilizaron son:

- **Matching por KNN (Pareo por N vecinos más cercanos):** para asignar a cada unidad no tratada una unidad tratada que sea lo más similar posible y viceversa. De esta manera contamos con un contra fáctico para cada unidad que haya recibido o no tratamiento. Se realizará este método con $N=1$ y con $N=5$.
- **Matching por KNN utilizando el puntaje de propensión:** en este caso no se realiza el matching considerando todas las variables observacionales que componen el set de datos, sino que se calcula un puntaje de propensión calculado mediante regresión logística y una técnica de machine leaning conocida como ‘boosting’, la cuál es es un técnica de aprendizaje supervisado con el objetivo de predecir la probabilidad de tratamiento. El puntaje de propensión tiene como ventaja la reducción de dimensionalidad de las variables observacionales a una sola. Nuevamente se realizará con $N=1$ y $N=5$.

- **Weighted Propensity Score Matching (Pareo ponderado por puntaje de propensión):** Una vez hallado el puntaje de propensión se ponderan de los resultados según el mismo. El efecto de tratamiento promedio se obtiene de la media del resultado ponderado.

Los tres procedimientos se aplican a las dos variables de respuesta sobre las que se trabaja: Resultados en Lengua y en Matemática.

Con el objetivo de obtener una distribución de los efectos estimados para cada técnica, se utiliza bootstrapping como esquema de resampling. Cada una de las técnicas se corrió veinte (20) veces tomando un 30% de los datos con reemplazo. La razón de esta decisión es poder generar la suficiente cantidad de corridas distintas que nos permitan estimar un intervalo de resultados posibles mediante los percentiles de la distribución de resultados.

6.3 Cálculo del Propensity Score

Se utilizaron dos técnicas distintas para el cálculo del Propensity Score. La primera es la regresión logística, la cuál es popular en este tipo de estudios (Stuart, 2007), y la segunda es el uso de técnicas de 'boosting'. Los Generalized Boosted Models (GBM) son técnicas de aprendizaje supervisado en las que algoritmos automáticos se adaptan a los datos y ajustan muchos árboles de decisión a los datos para mejorar la precisión de la predicción. Su adaptabilidad los hace atractivos para utilizar en modelos compuestos por una gran cantidad de variables para hacer predicciones de la variable tratamiento.¹⁰ En este estudio se utilizarán ambas técnicas a fin de aplicar tanto los métodos más tradicionales como otros más avanzados en la predicción del propensity score y así lograr un análisis más completo de la pregunta a resolver.

6.4 Controles y Supuestos

Como se ha mencionado anteriormente, la técnica de propensity score matching es tan buena como las variables observacionales que utiliza. Además de esto, hay algunas cuestiones que hay que tener en cuenta en la aplicación de la técnica. Una de ellas es la condición de soporte común, esto significa que debe haber observaciones de ambos grupos (tratamiento y control) en el rango de puntos de propensión a comparar. Los individuos que caen fuera del área común deben ser descartados ya que no cuentan con un miembro del otro grupo a comparar. Al mismo tiempo implica que el propensity score $P(x)$ no debe ser 0 ni 1. Lo que exige que ninguna observación pueda predecirse perfectamente ya que así asegura que

¹⁰ (McCaffrey, Ridgeway, & Morral, Propensity Score Estimation With Boosted Regression for Evaluating Causal Effects in Observational Studies, 2004)

otras observaciones con las mismas variables observables tengan probabilidad de estar en cualquiera de ambos grupos.

En los siguientes gráficos observamos el grado de overlap en el soporte de ambos grupos

Figura 14: Distribución de los casos de control y tratamiento en el puntaje de propensión previo al matching calculado por regresión logística.

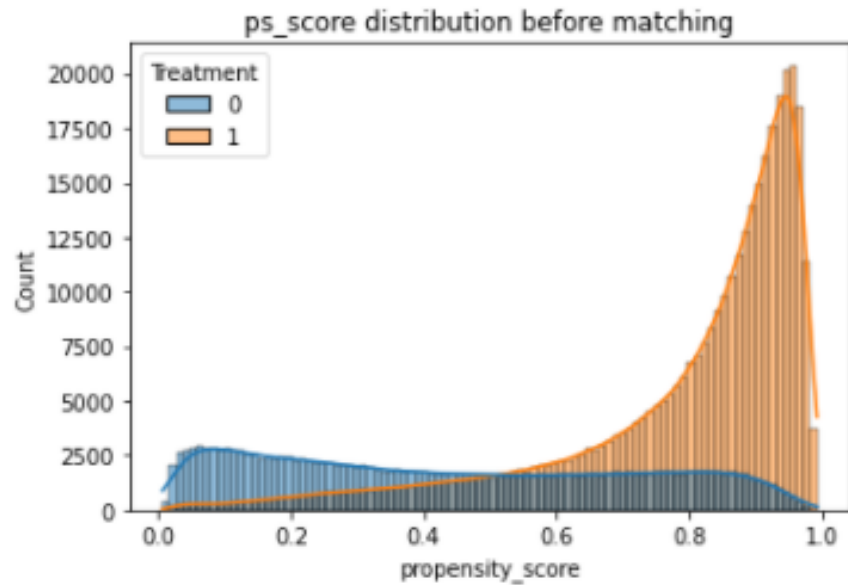
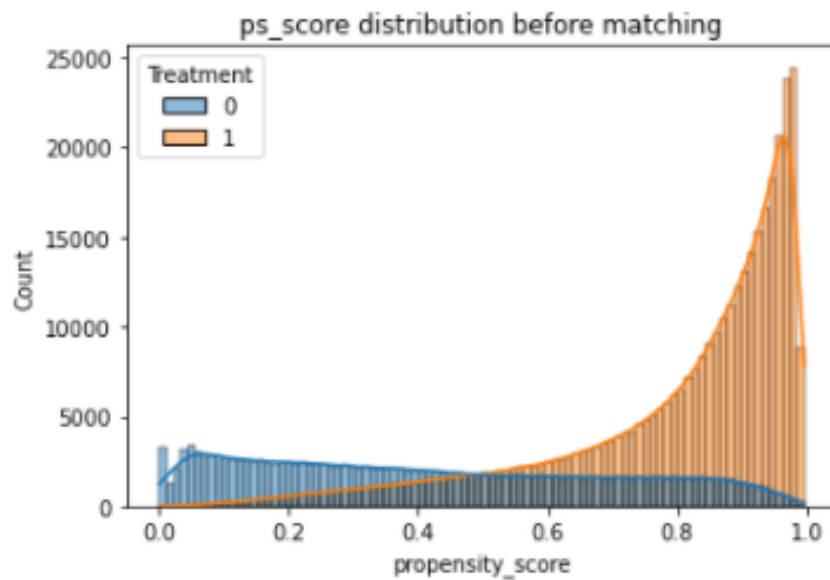


Figura 15: Distribución de los casos de control y tratamiento en el puntaje de propensión previo al matching calculado por boosting.



Se puede observar en ambos casos que contamos con observaciones a lo largo de todo el espacio del propensity score. Para el cumplimiento riguroso de esta condición se calcularon los mínimos y máximos puntajes de propensión para las observaciones de ambos grupos.

Para el propensity score calculado por regresión logística se observa que hay solapamiento de casos cuando el propensity score está entre 0,078 y 0,989, mientras que para boosting el solapamiento se da entre 0,0032 y 0,0992.

Se confirma que no hay observaciones con predicción 1 o 0 y al mismo tiempo se descartan las observaciones con $P(x) < 0.0078$ y las que tienen $P(x) > 0.989$ en el caso de la regresión logística, y con $P(x) < 0.0032$ y las que tienen $P(x) > 0.992$ en el caso de boosting, de forma que en todo el espacio del propensity score se encuentren tanto casos de tratamiento como de control.

Una vez hecha esta primera verificación, se pondrá el foco en otro de los supuestos: la independencia condicional. Esta consiste en que luego de hacer matching por el propensity score, la asignación a un grupo u otro sea “casi aleatoria”.

Este punto lo podemos analizar mediante la visualización de la distribución a lo largo del propensity score tras haber hecho el matching. El supuesto es que la chance de obtener una muestra de uno u otro grupo tras haber controlado por el propensity score sea similar.

Figura 16: Distribución de los casos de control y tratamiento en el puntaje de propensión calculado mediante regresión logística previo al matching

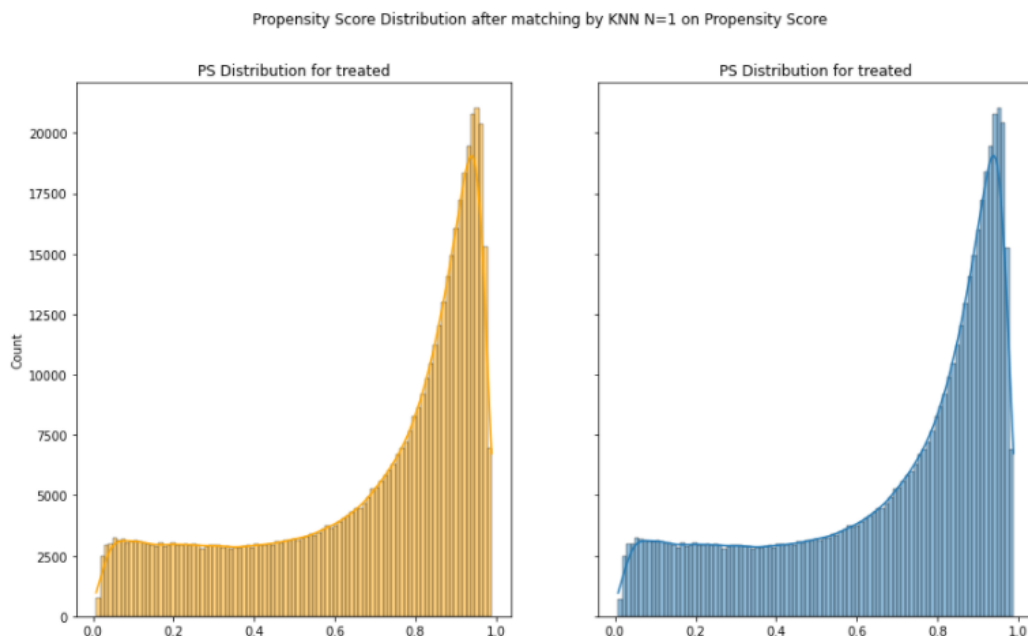
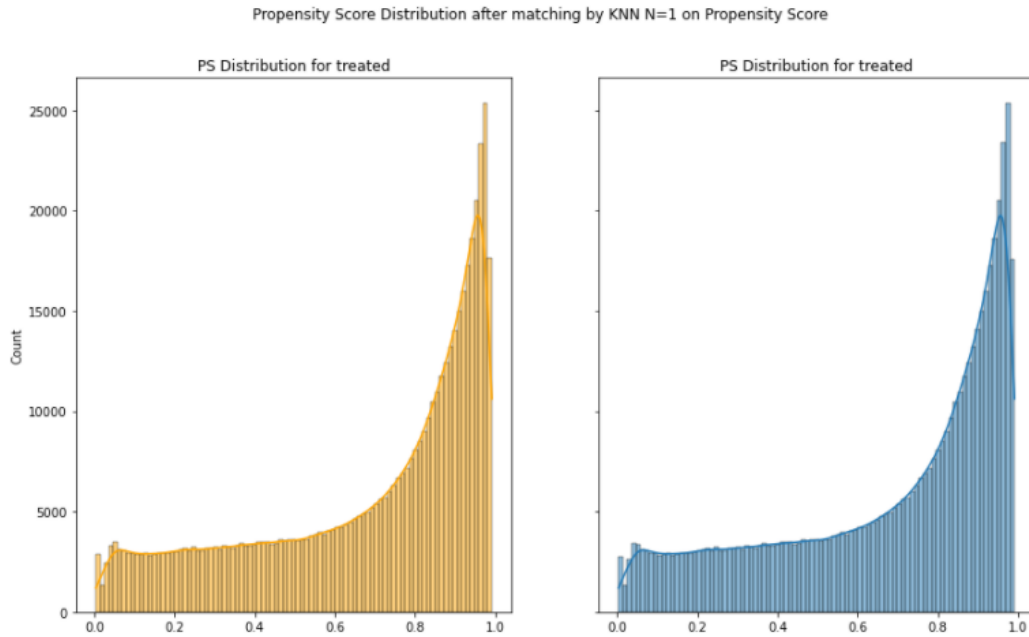


Figura 17: Distribución de los casos de control y tratamiento en el puntaje de propensión calculado mediante boosting previo al matching



El resultado es una mejora considerable respecto a la distribución previo al matching con knn1.

6.5 Resultados

Tabla 2: ATE estimado para KNN de variables

Estimación del ATE según técnica		
	Matching X (KNN N=1)	Matching X (KNN N=5)
Lengua	Mínimo	0,066
	Percentil 1%	0,067
	Media	0,082
	Percentil 99%	0,1
	Máximo	0,103
Matemática	Mínimo	0,06
	Percentil 1%	0,061
	Media	0,074
	Percentil 99%	0,091
	Máximo	0,091

Tabla 3: ATE estimado con técnicas de Propensity Score calculado mediante Regresión Logística

Estimación del ATE según técnica (PS Regresión Logística)

		Matching PS RL(KNN N=1)	Matching PS (KNN N=5)	Weighted Propensity
Lengua	Mínimo	0,032	0,029	0,029
	Percentil 1%	0,033	0,031	0,032
	Media	0,049	0,049	0,052
	Percentil 99%	0,063	0,063	0,062
	Máximo	0,064	0,064	0,063
Matemática	Mínimo	0,028	0,035	0,05
	Percentil 1%	0,029	0,036	0,05
	Media	0,051	0,049	0,06
	Percentil 99%	0,077	0,072	0,076
	Máximo	0,08	0,074	0,077

Tabla 4: ATE estimado con técnicas de Propensity Score calculado mediante Boosting

Estimación del ATE según técnica

		Matching PS (KNN N=1)	Matching PS (KNN N=5)	Weighted Propensity
Lengua	Mínimo	0,013	0,011	0,075
	Percentil 1%	0,014	0,012	0,076
	Media	0,032	0,03	0,089
	Percentil 99%	0,05	0,049	0,104
	Máximo	0,05	0,049	0,105
Matemática	Mínimo	0,014	0,017	0,072
	Percentil 1%	0,014	0,017	0,072
	Media	0,032	0,033	0,084
	Percentil 99%	0,05	0,055	0,097
	Máximo	0,05	0,057	0,097

Los resultados obtenidos para el ATE son similares más allá de técnicas de matching utilizadas y de los algoritmos utilizados para el cálculo del propensity score. Existe cierta diferencia entre los rangos en los que se distribuyeron los ATEs de matching mediante todas las variables observacionales y los obtenidos mediante el propensity score.

Figura 18: KDE de los ATE estimados en lengua según técnica de investigación causal aplicada. PS calculado con Regresión Logística

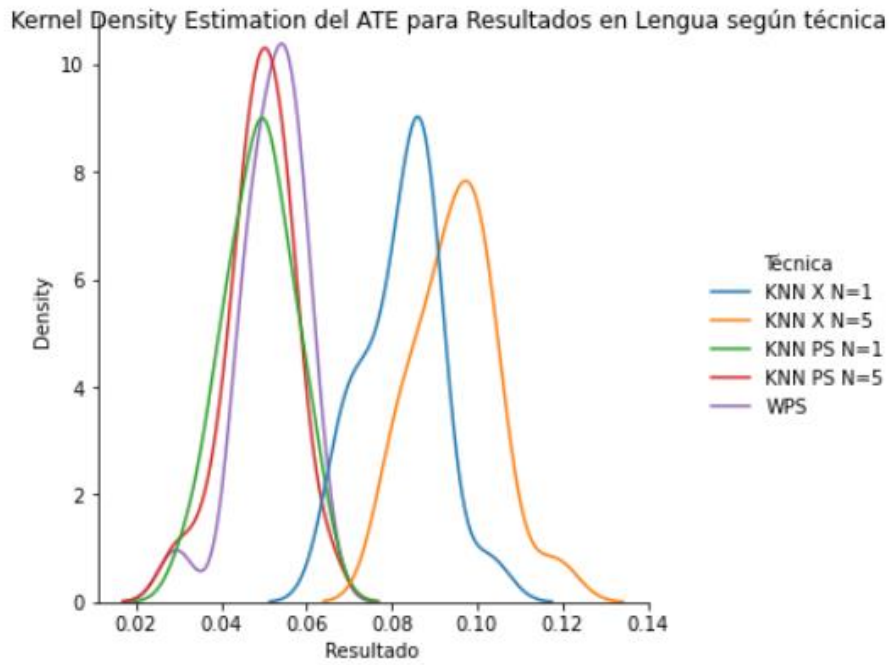


Figura 19: KDE de los ATE estimados en lengua según técnica de investigación causal aplicada. PS calculado con Boosting

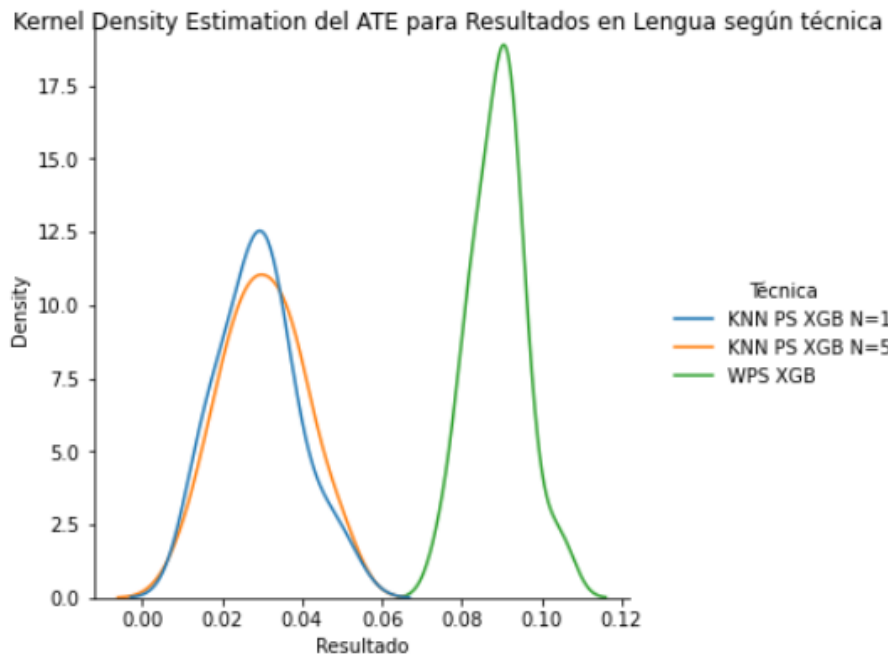


Figura 20: KDE de los ATE estimados en matemática según técnica de investigación causal aplicada PS calculado con Regresión Logística

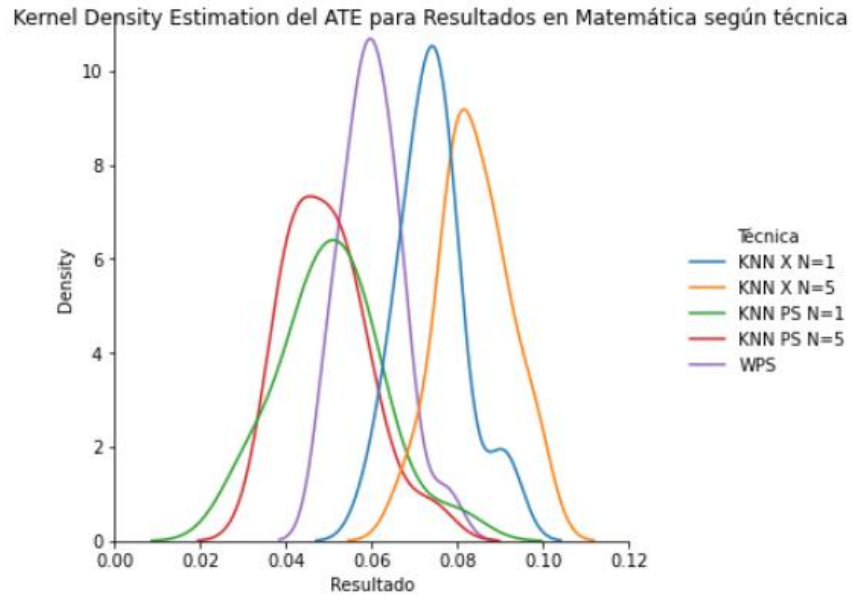
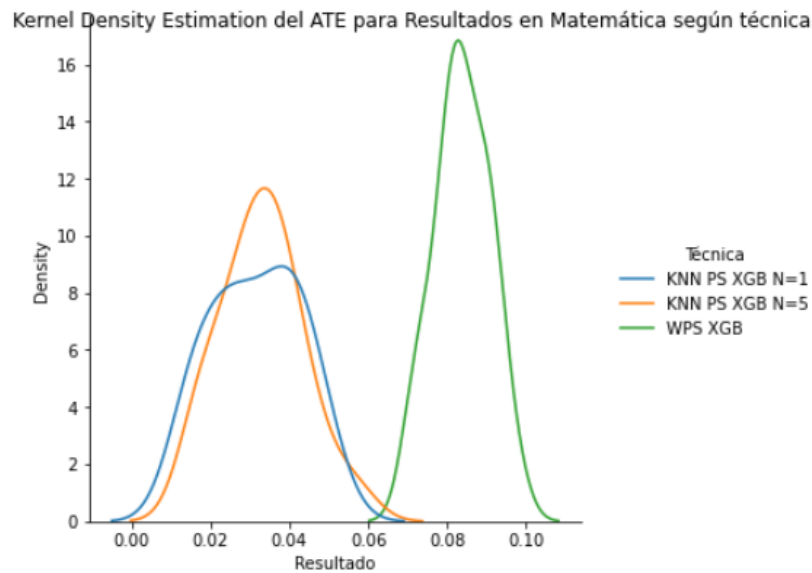


Figura 21: KDE de los ATE estimados en matemática según técnica de investigación causal aplicada. PS calculado con Boosting



La representación gráfica de los resultados muestra que los ATEs se concentran en la región entre 0 y 0,1 desvíos estándar. Aun cambiando la forma de calcular el propensity score, se puede observar que no ha habido grandes diferencias y que los resultados han sido consistentes.

El efecto promedio de tratamiento calculado es positivo para todas las técnicas. Para los resultados en lengua el ATE estimado varía entre 0,011 a 0,0118 según la técnica. En el caso de matemática los resultados hallados van de 0,014 a 0,099 desvíos estándar. En ninguno de los casos se halló un efecto negativo.

Tras haber hallado el impacto que tiene el tratamiento en cuestión, resta poner el foco en la pregunta que surge casi instintivamente: **¿Vale la pena invertir en la provisión gratuita de computadoras?** En la sección 7 exploraremos un posible marco para el análisis de costo-beneficio. Para un tratamiento detallado de los ejercicios de costo-beneficio ver (Boardman, Greenberg, Vining, & Weimer, 2018).

6.6 Corrección del sesgo generado por missings en la variable de respuesta

Como chequeo adicional de los resultados se corrige un potencial sesgo de selección de la variable de respuesta causada por los missings y obtuvimos resultados similares a los ejercicios anteriores.

La siguiente tabla muestra los resultados de la corrección, ponderando la respuesta por la probabilidad estimada de que esté completa la variable puntaje bajo el supuesto de MAR.

Tabla 5: Estimación del ATE corrección por posible sesgo de selección en variable respuesta

Estimación del ATE según técnica		
		Inversed Weighted
Lengua	Mínimo	0,018
	Percentil 1%	0,02
	Media	0,037
	Percentil 99%	0,051
	Máximo	0,051
Matemática	Mínimo	0,032
	Percentil 1%	0,032
	Media	0,045
	Percentil 99%	0,058
	Máximo	0,058

Figura 22: Kernel Density Estimation del ATE corrección por posible sesgo de selección en Resultados en Lengua

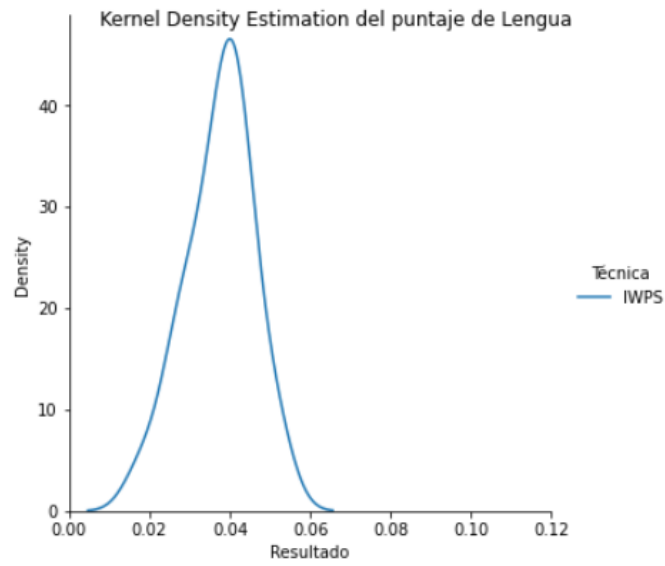
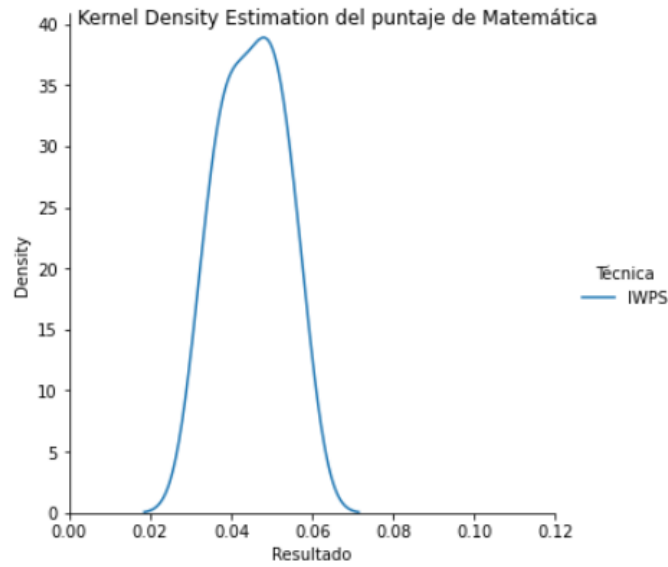


Figura 23: Kernel Density Estimation del ATE corrección por posible sesgo de selección para Resultados en Matemática



7. Marco de Referencia para Análisis de Efectividad del Costo

Para poder brindar una recomendación, a la evaluación de impacto debemos agregarle un análisis de costos. El análisis de costos consiste en comparar los impactos y costos de varios programas que buscan lograr el mismo resultado y así brindar solución a dos temas importantes en la evaluación de proyectos. El primero, es poder resumir un programa complejo en proporciones de efectos a costos, y el segundo, poder usar esta medida común para poder comparar múltiples programas que podrían tener origen en contextos y períodos diferentes. Para desarrollar la temática se utilizará de marco de referencia el informe desarrollado por el artículo *Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education* (Dhaliwal, Duflo, Glennerster, & Tulloch, 2011).

Existen dos maneras de abordar la temática, mediante la efectividad del costo o mediante el costo-beneficio.

- **Efectividad del Costo:** El análisis de efectividad del costo muestra la cantidad de “efecto” que un programa logra en un resultado a un costo dado.
- **Costo-Beneficio:** Combina los diferentes beneficios de un programa en una sola escala, la cual suele ser monetaria, y muestra la proporción de beneficios en relación con el costo.

La ventaja del costo-beneficio es que permite evaluar programas con múltiples resultados, y al poner los beneficios y los costos en una misma escala, el dinero, da mayor claridad en si un programa vale la pena la inversión, y al mismo tiempo permite comparar cual de todas las opciones genera el mejor retorno. Usualmente se utiliza para los programas que involucran una inversión inicial como la compra de una máquina, y luego se prorratan los beneficios que trae esta inversión a lo largo del tiempo. La desventaja es que requiere muchos supuestos acerca del valor monetario de los beneficios que no siempre están disponibles. Un ejemplo de esto son los referidos a calidad de vida; es muy subjetivo el valor que puede tener para una persona las consecuencias de tener que utilizar un bastón de por vida o tener que evitar cierto tipo de actividades. En el caso educativo, que es el que nos interesa en este estudio, un análisis de costo-beneficio requeriría que estimemos el aumento de productividad que traería incrementar el rendimiento escolar, sin embargo, no hay demasiado consenso en el valor monetario que pueda tener una mejora en los resultados de los exámenes, por lo que se dificulta esta tarea.

Por otro lado, el análisis de efectividad del costo es una alternativa que puede resultar útil si se determina un único resultado como objetivo de diferentes programas, como podría ser en este estudio el resultado en exámenes de lengua y matemática.

7.1 Cuantificando el impacto

Al comparar distintos programas posibles hay que tener en cuenta que los análisis de estos pueden provenir de diferentes estudios que se realizaron en países distintos y en épocas distintas. Es por ello por lo que es necesario estandarizar tanto los costos como los resultados para poder hacerlos comparables, lo que implica aplicar una misma metodología para ambos y también realizar ajustes que reflejen la diferencia temporal, de moneda y de inflación.

Otro de los detalles para tener en cuenta, es que es posible que un tratamiento tenga más de un efecto, lo cual no es fácil de captar con análisis de efectividad del costo ya que este está pensado para analizar solamente un resultado. Sin embargo, los análisis de costo-beneficio, que pueden ser más apropiados para evaluar más de un resultado, lo hacen al costo de transparencia ya que no es fácil ponerle un valor a cada uno de los efectos. Una opción es separar los efectos y asignarle al objetivo una porción correspondiente del costo y así poder continuar utilizando análisis de efectividad del costo. Un caso de ejemplo es el de servir alimentos en los colegios, por un lado, tiene como efecto aumentar la asistencia de los alumnos de menores recursos y por otro, también provee de una buena alimentación a esa población, lo cual es un efecto que podría no ser el que se está evaluando en esa ocasión y tiene repercusión en otros indicadores que podrían no estar relacionados a la educación directamente, como la salud. Nuevamente, la forma de evaluarlo depende del tomador de decisión y del objetivo que se persiga.

Al comparar estudios de impacto, es necesario tener presente la precisión que tiene cada uno. Lo primero que hay que tener en cuenta es el nivel de significancia que va a ser requerido al hacer el análisis de efectividad de costo, ya que tras definirlo al comparar diferentes tratamientos puede haber dos formas de insignificancia. La primera es que el resultado sea preciso pero muy cercano a cero, la segunda es que sea impreciso y los resultados varíen mucho, por lo cual no se pueda estimar su valor. Esto se manifiesta en mayor complejidad para los intervalos de confianza amplios que incluyen valores positivos y negativos, por lo que se recomienda excluirlos. Es posible estudiar la sensibilidad de la efectividad del costo al evaluar

los diferentes programas, no sólo por el resultado medio sino también por su cuartil inferior y superior para poder observar si al hacer un ranking entre programas varían mucho de posición.

Un factor importante para tener en cuenta es el “efecto derrame” de un tratamiento. Un ejemplo claro para describirlo es en un tratamiento por vacunación. Si se analiza la efectividad del costo de comprar vacunas, también habrá que tener en cuenta el efecto derrame de que las personas inoculadas potencialmente dejen de ser vectores de transmisión de la enfermedad y así repercutirá también en una baja de casos en la población de control. El efecto derrame solo debe ser tenido en cuenta cuando el tratamiento se hace a gran escala y se tiene debidamente medido el derrame.

7.2 Cuantificando costos

Para el cálculo de los costos de un programa es necesario considerar no solo la inversión que se realiza, sino también dentro de que período temporal. La información relacionada a los costos de un programa evaluado no suele ser completa, por lo que usualmente es necesario un proceso de recopilación de información consultando a las fuentes primarias para poder especificar todos los “ingredientes” necesarios para el tratamiento y también el costo de cada uno de ellos. A su vez, tener en cuenta el marco donde se realizó el estudio y su similitud con donde se evalúa aplicarlo nuevamente, ya que existe la posibilidad de que, además del costo unitario por tratamiento, exista también la necesidad de una inversión en infraestructura para poder llevarlo a cabo. Es necesario tener presente quién sería el usuario del tratamiento para poder estimar correctamente los costos.

Se define la efectividad del costo de un programa como el cambio marginal, en nuestro caso de puntajes en lengua y matemática, dividido por el cambio marginal en costos por implementar el programa.

$$\text{Ratio Beneficio Costo} = \frac{[Puntajes con el programa] - [Puntajes sin el programa]}{[Costos con el programa] - [Costos sin el programa]}$$

De esta manera se pueden ignorar factores que ya existen sin el programa, como podría ser la infraestructura escolar si esta fuera suficiente para el programa, e incluir los costos relacionados al tratamiento como podría ser el valor de la computadora. Es por ello por lo que debe tenerse muy en cuenta el ámbito donde se aplicaría el tratamiento. Esto implica también considerar si se analiza el programa desde una institución o desde la sociedad en general. Si hubiera una donación de algún insumo

este debería ser considerado en los costos si se evalúa desde el punto de vista de la sociedad entera, pero si se analiza desde una institución particular entonces el costo de estos sería nulo.

Los costos “ocultos” a los beneficiarios también deberían ser tomados en cuenta. Estos pueden ser costos de tiempo, traslados u otros requerimientos que tenga el programa, ya que si bien no aplican a quien está implementando el programa, son un costo real para el usuario y deberían ser considerados. Su inclusión en la evaluación de efectividad de costos puede tener un impacto significativo en el resultado final. Particularmente este punto impacta cuando se evalúa un programa desde el punto de vista de una institución o desde la sociedad. Los costos de traslado de un bien deben ser considerados cuando se evalúa el programa desde una institución ya que debe invertirse dinero en ello. Al evaluarlo desde la sociedad estos costos no se incluyen ya que se consideran simplemente una redistribución del dinero que sigue quedando en la sociedad. La inclusión o no de estos costos es a juicio del tomador de decisiones de acuerdo con el ámbito en el que se aplica el programa.

Existen también los costos de escalabilidad, los cuales se manifiestan en la relación que hay entre costos fijos y variables en la etapa experimental de programa contra una etapa de gran escala. Existe la posibilidad de que los costos fijos sean muy altos como para hacer viable un programa a pequeña escala, pero al escalarlo se balancea por la gran cantidad de beneficiarios. Esto también impacta en los costos variables ya que la compra de insumos en gran volumen permite una disminución del costo unitario. Sin embargo, Dhaliwal et al. recomiendan mantener los costos y beneficios del programa como fueron analizados ya que la economía de escala puede traer otras complejidades como disminución del impacto al bajar la calidad del control sobre el programa.

7.3 Ajustes de unidades

Al estar comparando proyectos evaluados en diferentes momentos y países, es necesario traer los costos y beneficios a unidades comunes, ya sea ajustando por inflación, por monedas o por año de implementación.

Es necesario definir un año de análisis, el actual, para evaluar todos los programas y así evitar diferencias en costos que podrían ser causadas por la inflación. Lo mismo ocurre cuando los costos e impactos están distribuidos a través del tiempo, por lo que es necesario llevarlos primero al valor presente en el inicio del

programa evaluado y luego llevado al año actual de análisis. La tasa de descuento a aplicar varía según el país, pero queda a criterio del tomador de decisión elegir la tasa a utilizar.

Para el caso de los beneficios es más complejo ya que el beneficio puede no ser dinero, por lo que habría que definir una “tasa social de preferencia temporal”. Esta equivale a la tasa de unidades que estaría dispuesto el beneficiario a consumir el día de mañana a cambio de no consumir hoy. Esto es muy sensible al beneficio en cuestión y aun así es difícil de estimar. En los casos en que no hay un tiempo largo entre la implementación del programa y el beneficio, no hay tanta sensibilidad a la tasa mencionada. Sí sería razonable tenerla en cuenta cuando se requiere una inversión grande al comienzo y los beneficios se reciben después de unos años.

Para el caso de la inflación, cuando se trata de proyectos que requieren inversiones a lo largo de los años, los costos deberían llevarse al año base que se implementó el proyecto y luego ese valor total debe ser llevado al año de análisis que sería el actual.

7.4 Consideraciones Generales

Las evaluaciones de impacto proveen una imagen estática del efecto que tiene un tratamiento. Algunos de ellos están diseñados para medir el impacto de corto plazo, mientras que otros el de largo. Aun cuando se tome uno de largo plazo, hay que considerar la posibilidad de que las personas cambien su respuesta al programa una vez que esté más establecido. En algunos casos la respuesta es cada vez más positiva, mientras que en otros se obtienen retornos decrecientes como sería en el caso de repartir vacunas. Esto es muy dependiente del estado inicial del que se parte, y en que punto de la curva de beneficio marginal se encuentra. En India se utilizó un tratamiento antiparasitario en escuelas como incentivo para aumentar la asistencia en regiones donde había gran cantidad de casos de infección por gusanos intestinales. Si bien el tratamiento contra la infección es igualmente efectivo en Asia o en América, existe un beneficio marginal decreciente de seguir con el tratamiento, por lo que los beneficios obtenidos en regiones donde no prolifera la enfermedad van a ser menores.

Es por ello que es necesario tener en cuenta el punto de partida del programa tanto en la evaluación observada como en el ámbito donde se estudia aplicar. El objetivo es estimar lo más correctamente posible el costo-beneficio de los diferentes tratamientos, teniendo en cuenta que los resultados en los estudios de evaluación de impacto sólo son representativos en el contexto particular donde se hicieron.

Lo mismo aplica a los costos, en algunas regiones quizás sea más difícil tener acceso a los insumos necesarios para llevar a cabo el tratamiento, por lo que es recomendable utilizar estudios que hayan sido llevado a cabo en condiciones similares.

El estudio de efectividad de costo sirve para brindar información a la discusión de cuál es la mejor forma de mejorar un resultado. Son muchas las consideraciones para tener en cuenta para poder llevar los resultados de estudios realizados en otros países al lugar y tiempo donde se evalúa aplicar. Es un marco valioso y útil para mejorar la toma de decisiones en ámbitos que no son tan fáciles de cuantificar como lo es la educación o la salud. Con este tipo de análisis se puede proveer al tomador de decisión con herramientas que son claras en los supuestos y que pueden ser adaptadas para lograr mayor transparencia y menor subjetividad a la toma de decisión.

Centrándonos en el presente estudio, las experiencias del Banco Interamericano de Desarrollo (Cristia, 2012) y del experimento de entrega de computadoras a alumnos de sexto grado en California (Fairlie & Robinson, 2013) son las más apropiadas para extrapolar a la Argentina. El hecho de que se limiten a la entrega de computadoras y que abarque a alumnos de nivel primario y medio que no pueden acceder a un ordenador en el hogar se asemeja mucho al caso argentino. Por otro lado, se destaca que la experiencia israelí (Angrist & Lavy, 2002) no es comparable ya que en dicho estudio se buscó la completa integración de las computadoras en el aula, las cual involucró la capacitación de maestros, la creación de material didáctico adaptado y el uso de computadoras en clase. Es posible extrapolar los costos del programa del BID a Argentina, por las similitudes entre Perú y Argentina, y por el objetivo que persiguió dicho estudio. En ese caso cada computadora tuvo un costo aproximado de 200 dólares, si se entregara a todos los alumnos de sexto grado del set de datos (579.510 alumnos), se requeriría de una inversión de 116 millones de dólares aproximadamente. El presupuesto educativo para el año 2021 es de \$355.652 millones de pesos¹¹, lo que tomando una cotización del dólar de \$96, equivaldría a casi 3.700 millones de dólares. La inversión en las computadoras equivale al 3,13% del presupuesto anual en educación. Esta simple simulación del costo del programa sirve para poner en perspectiva el valor equivalente que deberían brindar los computadores en el rendimiento escolar. Dicho cálculo es complejo ya que habría que asignarle un valor monetario al potencial impacto, pero es apropiado tener en cuenta este balance de valores a la hora de impulsar una política pública de este sentido.

¹¹ (CIPPEC, 2021)

8. Conclusiones

La pregunta que nos propusimos responder fue: ¿Hay evidencia de efectos positivos sobre la performance en lengua y/o matemática de contar con una computadora en el hogar? En base a los datos públicos de las pruebas Aprender pudimos obtener algunas estimaciones del ATE poblacional. Esto nos llevó a la exploración de diferentes métodos de inferencia causal. A su vez, se obtuvieron estimaciones para cinco métodos distintos. Además, se aplicó *bootstrapping* para conocer la distribución de la estimación de ATE poblacional con cada una de las técnicas.

Los resultados están en línea con la literatura. (Fairlie & Robinson, 2013) y (Cristia, 2012) obtienen estimaciones de un orden de magnitud similar para el efecto del acceso a computadoras en el hogar. En nuestro caso el ATE se encontró entre 0,012 a 0,116 desvíos estándar para los resultados en lengua al tomar los percentiles 1 y 99 como referencia y de 0,014 a 0,098 en matemática para el mismo intervalo de percentiles.

El análisis de contar con una computadora en el hogar es valioso ya que permitió determinar que el efecto es positivo, aunque de una magnitud pequeña. Al mismo tiempo se presenta la metodología desarrollada por (Dhaliwal, Duflo, Glennerster, & Tulloch, 2011) de realizar ejercicios de costo beneficio para la evaluación de políticas públicas. Este marco metodológico provee al tomador de decisión con las herramientas y supuestos claros para evaluar si la distribución de computadoras a los alumnos es la medida más apropiada para mejorar su rendimiento académico, o si, por el contrario, otro tratamiento investigado en otro estudio de impacto pudiera tener mayor efectividad una vez que se toman en cuenta los costos. Como chequeo adicional y para contemplar un posible sesgo de selección se realiza una corrección por los missings en la estimación del ATE. En este caso, bajo el supuesto de MAR, consideraremos los missings como parte del mecanismo de selección muestral. Siguiendo la tradición del inverse probability weighting method (IPW) propuesto por (Horvitz & Thompson, 1952) ponderamos por la probabilidad de tener completo el puntaje del examen condicional a los controles (X) bajo el supuesto del cumplimiento del supuesto de MAR. Los resultados obtenidos son similares a los ejercicios anteriores en cuanto al impacto del tratamiento, tanto en los puntajes de lengua como de matemática.

9. Posibles Extensiones

(Hernán & Robins, 2020) explican que perseguir la predicción más exacta posible de la variable tratamiento puede ser contraproducente, ya que es posible que para lograr la mayor exactitud se incluyan variables que agreguen mayor sesgo al modelo. Cita como ejemplo ilustrativo que, si en un estudio de un tratamiento médico existiera la variable hospital, en la cual en el hospital A se realiza el tratamiento al 99% de los enfermos, mientras que en el B sólo al 1%. Al buscar la mayor exactitud posible en la predicción de la posibilidad de tratamiento, tendría sentido incluir esta variable, pero al tomar una población al azar existirá la chance de que todos los que fueron al hospital A hayan recibido tratamiento o en el caso del hospital B de que nadie lo haya recibido. En ese caso tendríamos un propensity score de 1 absoluto (o 0 absoluto en el caso inverso), lo hace imposible determinar el efecto causal para estos individuos ya que no será posible encontrar grupo de comparación para ellos.

Existen experiencias como la de (McCandless, Gustafson, & Austin, 2009), en la que utilizan modelos bayesianos mediante el método MCMC (Markov Chain Monte Carlo) para modelar la incertidumbre en el propensity score. Una característica del modelo aplicado es que ajusta tanto para el tratamiento como para el resultado simultáneamente, en vez de calcular primero el tratamiento y luego el resultado. Esta característica tiene como debilidad que aumenta el impacto de un posible error de diseño del modelo. Los autores mencionan que hay diferentes puntos de vista en la literatura acerca de asistir con la variable resultado a la estimación del propensity score. Quienes se oponen argumentan que no contar con la variable desenlace reproduce el escenario de una prueba aleatoria. Sin embargo, el modelo bayesiano mencionado reproduce con mayor fidelidad el verdadero propensity score, por lo que se recomienda al lector interesado expandir el presente trabajo con la utilización de este tipo de técnicas.

Por otro lado, (Westreich, Lessler, & Jonsson Funk, 2010) evaluaron el uso de técnicas alternativas a la regresión logística para el cálculo del propensity score, entre ellas las redes neuronales, support vector machines (SVM), árboles de decisión y boosting. En su estudio, los autores comentan que de las alternativas que evaluaron, boosting es la que mejor potencial tiene debido a que no requieren tanta experiencia en la selección de hiperparámetros como ocurre en redes neuronales o SVM. Boosting tiene la ventaja de tener una “implementación ingenua”, lo que facilita su aplicación. Los autores argumentan que las técnicas de machine learning hacen menos supuestos que la regresión logística, la cual es paramétrica. Esto les da mayor flexibilidad para interactuar con no linealidades, lo que les genera optimismo en que puedan llegar a reemplazar a la regresión logística en el cálculo de propensity score.

Si bien en el presente estudio se utilizaron tanto la regresión logística como boosting para la estimación del impacto de acceso a computadoras en el resultado académico en Argentina, se aconseja explorar la utilización de los otros métodos mencionados como redes neuronales, SVM o árboles de decisión.

10. Bibliografía

- Alves, M. F. (2020). *Causal Inference for The Brave and True*.
- Angrist, J., & Lavy, V. (2002). New Evidence On Classroom Computers and Pupil Learning. *The Economic Journal*, 735-765.
- Angrist, J., & Pischke, J. S. (2014). *Mastering 'Metrics: The path from Cause to Effect*. Princeton University Press.
- Berrueta-Clement, J. (1984). *Changed Lives: The Effects of the Perry Preschool Program on Youths through Age 19*. Ypsilanti, Mich.: High/Scope Educational Research Foundation,.
- Boardman, A., Greenberg, D., Vining, A., & Weimer, D. (2018). *Cost-Benefit Analysis Concepts and Practice*. Cambridge: Cambridge University Press.
- Branson, N., Hofmeyr, C., & Lam, D. (2014). Progress through school and the determinants of school dropout in South Africa. *Development Southern Africa*, 106-126.
- Brookhart, A., Schneeweiss, S., Rothman, K., Glynn, R., Avorn, J., & Stürmer, T. (2006). Variable Selection for Propensity Score Models. *American Journal of Epidemiology*, 1149-1156.
- Byrne, B. M. (1984). The General/Academic Self-Concept Nomological Network: A Review of Construct Validation Research. *Review of Educational Research* , 427-456.
- Carolan, B., & Wasserman, S. (2014). Does parenting style matter? Concerted cultivation, educational expectations, and the transmission of educational advantage. *Sociological Perspectives*, 168-186.
- Catterall, J. S. (1987). On the Social Costs of Dropping out of School. *The High School Journal*, 19-30.
- CIPPEC. (19 de Julio de 2021). *Educación en tiempos de pandemia: el gobierno incrementa la inversión educativa*. Obtenido de CIPPEC: <https://www.cippec.org/textual/educar-en-tiempos-de-pandemia-el-gobierno-incrementa-la-inversion-educativa/>
- Cordero, J., & Cristóbal, V. (2017). CAUSAL INFERENCE ON EDUCATION POLICIES: A SURVEY OF EMPIRICAL STUDIES USING PISA, TIMSS AND PIRLS. *Journal of Economic Surveys*, 1-38.
- Cristia, J. (2012). *Technology and Child Development: Evidence from the One Laptop per Child Program*. Washington DC: IDB Working Paper IDB-WP-304.
- Davis-Kean, P. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology*, 294-304.
- Dhaliwal, I., Duflo, E., Glennerster, R., & Tulloch, C. (2011). Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries: A General Framework with Applications for Education. *Abdul Latif Jameel Poverty Action Lab*, 1-50.
- Elwert, F. (2013). Graphical Causal Models. *Handbook of Causal Analysis for Social Research*, 978-994.

- Fairlie, R., & Robinson, J. (2013). Experimental Evidence on the Effects of Home Computers on Academic Achievement among Schoolchildren. *American Economic Journal: Applied Economics*, 211-240.
- Fan, X., & Nowell, D. (2011). Using Propensity Score Matching in Educational Research. *Gifted Child Quarterly*, 74-79.
- Farley, D. O. (2002). Effects of CAHPS Health Plan Performance Information on Plan Choices by New Jersey Medicaid Beneficiaries. *HSR: Health Services Research*, 985-1007.
- Finn, J. D. (1989). Withdrawing From School. *Review of Educational Research*, 117-142.
- Finn, J., & Gerber, S. (2005). Small Classes in the Early Grades, Academic Achievement, and Graduating From High School. *American Psychological Association*, 214-223.
- Galiani, S. (21 de Mayo de 2021). *Slide Player*. Obtenido de Impact Evaluation Sebastian Galiani November 2006 Matching Techniques.: <https://slideplayer.com/slide/8741140/>
- Gelman, A. (2011). Review Essay Causality and Statistical Learning. *American Journal of Sociology*, 955-966.
- Gertler, P., Martínez, S., Premand, P., Rawlings, L., & Vermeersch, C. (2017). *La evaluación de impacto en la práctica*. Washington DC: Grupo Banco Mundial.
- Gustafsson, J.-E. (2008). Gustafsson, J.-E. (2008). Effects of International Comparative Studies on Educational Quality on the Quality of Educational Research. *European Educational Research Journal*, 1-17.
- Han, F. (2018). Estimation of the Average Treatment Effect with Missing Outcome Data . *Advances in Computer Science Research*, 377-382.
- Hernán, M., & Robins, J. (2020). *Causal Inference: What If*. Boca Raton: Chapman & Hall.
- Hillman, N. (31 de May de 2021). *Using difference-in-differences in higher education research*. Obtenido de Nick Hillman, Associate Professor, UW-Madison: <https://web.education.wisc.edu/nwhillman/index.php/2018/07/10/using-difference-in-differences-in-higher-education-research/>
- Horvitz, D. G., & Thompson, D. J. (1952). A Generalization of Sampling Without Replacement From a Finite Universe. *Journal of the American Statistical Association*, 663-685.
- Li, C. (2013). Little's test of missing completely at random. *The Stata Journal*, 795-809.
- Little, R. J. (1995). Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *Journal of the American Statistical Association*, 1112-1121.
- Lloyd, D. N. (1978). Prediction of School Failure From Third-Grade Data. *Educational and Psychological Measurement*, 1993-1200.
- Lovejoy, C. (13 de June de 2021). *From Correlation to Causation in Machine Learning: Why and How*. Obtenido de Towards Data Science: <https://towardsdatascience.com/from-correlation-to-causation-in-machine-learning-why-and-how-4485bca8d145>

- McCaffrey, D., Ridgeway, G., & Morral, A. (2004). Propensity Score Estimation With Boosted Regression for Evaluating Causal Effects in Observational Studies. *Psychological Methods*, 403-425.
- McCaffrey, D., Ridgeway, G., & Morral, A. (2004). Propensity Score Estimation With Boosted Regression for Evaluating Causal Effects in Observational Studies. *American Psychological Association*, 403-425.
- McCandless, L., Gustafson, P., & Austin, P. (2009). Bayesian propensity score analysis for observational data. *STATISTICS IN MEDICINE*, 94–112.
- McElreath, R. (31 de May de 2021). *Twitter*. Obtenido de Twitter: <https://twitter.com/rmcelreath/status/1100763647943557121>
- Ministerio de Educación. (19 de December de 2020). *Argentina.gob.ar*. Obtenido de Aprender. Evaluación e Información Educativa: <https://argentina.gob.ar/educacion/evaluacion-informacion-educativa/aprender>
- Ministerio de Educación de la República Argentina. (Enero de 2020). *Evaluación e Información Educativa: Anuarios Estadísticos*. Obtenido de Argentina.gob.ar: <https://www.argentina.gob.ar/educacion/evaluacion-informacion-educativa/anuarios>
- Ministerio de Educación, C. C. (2018). *Aprender 2018. Notas Técnicas*. Ministerio de Educación, Cultura, Ciencia y Tecnología.
- Ministerio de Educación, Cultura, Ciencia y Tecnología de la República Argentina. (2019). *Aprender 2018. Informe Nacional de Resultados 6 Año Nivel Primario*. Ciudad de Buenos Aires, República Argentina: Ministerio de Educación, Cultura, Ciencia y Tecnología de la República Argentina.
- Mohamed, S. (30 de May de 2021). *Machine Learning Trick of the Day (8): Instrumental Thinking*. Obtenido de The Spectator: <http://blog.shakirm.com/2018/10/machine-learning-trick-of-the-day-8-instrumental-thinking/>
- OECD. (2016). *"Immigrant background, student performance and students' attitudes towards science", in PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris: OECD Publishing.
- OECD. (2020). *"Girls' and boys' performance in PISA", in PISA 2018 Results (Volume II): Where All Students Can Succeed*. Paris: OECD Publishing.
- OECD. (2020). *Education at a Glance 2020: OECD Indicators*. Paris: OECD Publishing.
- Panel de indicadores*. (Enero de 2020). Obtenido de Argentinos por la Educación: https://argentinosporlaeducacion.org/indicador/primario_secundario_superior/8/Tasa%20de%20abandono%20intraanual/72qkub
- Pearl, J. (2009). Causal inference in statistics: An Overview. *Statistic Surveys*.
- Robins, J. M., Rotnitzky, A., & Zhao, L. P. (1995). Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *Journal of the American Statistical Association*.

- Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effect. *Biometrika*, 41-55.
- Rosenbaum, P., & Rubin, D. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association*, 516-524.
- Rubin, D. (2005). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association*, 322-331.
- Rubin, D., & Thomas, N. (1996). Matching Using Estimated Propensity Scores: Relating Theory to Practice. *International Biometric Society*, 249-264.
- Rumberger, R. W. (1987). High School Dropouts: A Review of Issues and Evidence. *Review of Educational Research*, 101-121.
- Rumberger, R., & Rotermund, S. (2012). The Relationship Between Engagement and High School Dropout. *Handbook of Research on Student Engagement*, 491-513.
- Stuart, E. (2007). Estimating Causal Effects Using School-Level Data Sets. *Educational Researcher*, 187-198.
- Westreich, D., Lessler, J., & Jonsson Funk, M. (2010). Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *Journal of Clinical Epidemiology*, 826-833.