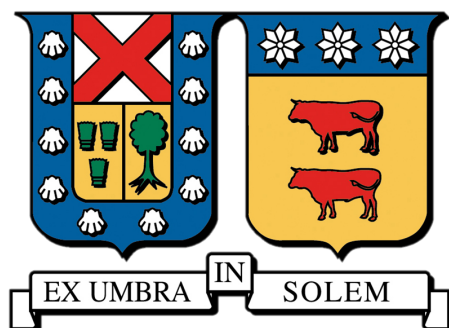




UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE INDUSTRIAS

**SISTEMA DE ADMISIÓN ESCOLAR EN CHILE: EVOLUCIÓN DE LA
SEGREGACIÓN SOCIOECONÓMICA, EL RENDIMIENTO
ACADÉMICO Y LA ASIGNACIÓN DE ESTUDIANTES TRAS SU
IMPLEMENTACIÓN**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

AUTOR

CONSTANZA SOFÍA BERRÍOS GUTIÉRREZ

PROFESOR GUÍA

BERNARDO PINCHEIRA SARMIENTO

PROFESOR CO-REFERENTE

RODOLFO SALAZAR ALBORNOZ

VALPARAÍSO DE CHILE, OCTUBRE 2026



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: SISTEMA DE ADMISIÓN ESCOLAR EN CHILE: EVOLUCIÓN DE LA SEGREGACIÓN SOCIOECONOMICA, EL RENDIMIENTO ACADÉMICO Y LA ASIGNACIÓN DE ESTUDIANTES TRAS SU IMPLEMENTACIÓN

Nombre del candidato(a): Constanza Berríos Gutiérrez

Carrera / Grado: Ingeniería Civil Industrial

Campus: Casa Central **Departamento:** Industrias

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Bernardo Pincheira, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente

DEJO CONSTANCIA que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 22-10-2026

Firma:

Estudiante o Candidato(a):

Fecha: 22-10-2026

Firma:

Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Tabla de Contenidos

1. Problema de Investigación	5
2. Objetivos	7
2.1. Objetivo General	7
2.2. Objetivos Específicos	7
3. Marco Teórico	8
3.1. Sistema Educacional en Chile	8
3.1.1. Niveles de educación	8
3.1.2. Tipos de financiamiento	9
3.2. Sistema de Admisión Escolar	11
3.3. Segregación Escolar	15
3.3.1. Elección de Establecimiento	15
3.4. Competencia entre establecimientos	16
3.4.1. El Sistema de Medición de la Calidad de la Educación (SIMCE)	18
3.5. Modelos Econométricos	18
3.5.1. Medición de la Segregación	19
3.5.2. Diferencia de Diferencias	21
4. Evolución de la segregación socioeconómica: Índice de Duncan	23
4.1. Descripción de Datos	23
4.1.1. Fuente de datos	23
4.1.2. Tablas de datos utilizadas	24

4.2.	Consolidación de Datos	27
4.2.1.	Carga de datos	28
4.2.2.	Fusión de las bases de datos	30
4.2.3.	Limpieza de datos	31
4.2.4.	Creación de variables	32
4.3.	Metodología	33
4.3.1.	Definición de grupos socioeconómicos	33
4.3.2.	Cálculo del Índice de Duncan	34
4.4.	Estadísticas Descriptivas	36
4.4.1.	Distribución socioeconómica de los estudiantes por región	36
4.4.2.	Distribución territorial y ruralidad	39
4.4.3.	Disponibilidad regional de establecimientos educacionales	42
4.5.	Resultados	45
5.	Efecto del SAE sobre el rendimiento académico: Puntajes SIMCE	58
5.1.	Descripción de Datos	58
5.1.1.	Fuente de datos	58
5.1.2.	Tablas de datos utilizadas	58
5.2.	Consolidación de Datos	59
5.2.1.	Fusión de las bases de datos	59
5.2.2.	Limpieza de datos	60
5.2.3.	Creación de variables	62
5.2.4.	Construcción de la base para el modelo DiD	63

5.3. Metodología	64
5.3.1. Variables del modelo	65
5.3.2. Fórmula del modelo	67
5.3.3. Descripción de los modelos	69
5.4. Estadísticas Descriptivas	71
5.4.1. Puntajes SIMCE de lectura y matemática	71
5.5. Resultados	72
5.5.1. Resultados para lectura	73
5.5.2. Resultados para matemática	76
6. Asignación escolar y primera preferencia en el Sistema de Admisión Escolar	82
6.1. Descripción de Datos	82
6.1.1. Fuente de datos	82
6.1.2. Tablas de datos	83
6.2. Consolidación de Datos	86
6.2.1. Fusión de las bases de datos	86
6.2.2. Construcción de la variable de asignación	87
6.2.3. Construcción de la información geográfica	88
6.3. Metodología	90
6.3.1. Análisis de asignación a primera preferencia	90
6.3.2. Análisis de distancia entre residencia y establecimientos	90
6.3.3. Estadísticas descriptivas	93
6.4. Resultados	96



6.4.1. Asignación al establecimiento de primera preferencia	96
6.4.2. Distancia entre el establecimiento asignado y la primera preferencia	98
7. Conclusiones	101
8. Limitaciones	105
9. Anexos	110
9.1. Anexo A	110
9.2. Anexo B	111
9.3. Anexo C	112

1. Problema de Investigación

En Chile, el Sistema de Admisión Escolar (SAE) fue creado con el propósito de reemplazar los procesos de selección discrecional por un mecanismo centralizado y transparente, reducir la segregación social y promover el acceso equitativo a la educación, especialmente para estudiantes más vulnerables.

Sin embargo, a pesar de su propósito inclusivo, diversos estudios y observaciones preliminares sugieren que el SAE no ha logrado reducir de manera efectiva la segregación socioeconómica. Familias con mayores recursos logran mantener ventajas en el proceso de postulación por encima de los estudiantes con menores recursos. Además, la preferencia de las familias de mayores ingresos por escuelas en las que sus hijos compartan con pares de características socioeconómicas similares continúa siendo un factor determinante de la segregación escolar (Gazmuri, 2024). Esta tendencia se evidencia en la migración de estudiantes desde establecimientos con financiamiento estatal hacia colegios privados (Kutscher et al., 2023).

Resulta igualmente relevante analizar la diferencia entre la asignación real y la asignación deseada por los padres que participan en este sistema de admisión. Dicha brecha puede verse fuertemente influida por la segregación residencial, la cual genera una distribución desigual de los estudiantes según su nivel socioeconómico (Santos & Elacqua, 2017). Esta problemática no solo se observa en Chile, sino también en otros países, como Inglaterra, donde se ha evidenciado que los estudiantes de mayor nivel socioeconómico tienen una probabilidad más alta de ser admitidos en escuelas de mejor calidad (Burgess et al., 2014).

Es por esto que resulta importante estudiar las reformas de asignación escolar descentralizadas, ya que, si bien la elección de escuela puede mejorar la eficiencia en las postulaciones,

también puede tener costos en términos de segregación tanto socioeconómica como académica (Greaves, 2024).

En base a lo mencionado surgen las preguntas: ¿En qué medida el Sistema de Admisión Escolar (SAE) ha contribuido a disminuir o, por el contrario, a mantener o profundizar la segregación socioeconómica en Chile, considerando desde el año en que se implementó hasta 2024? ¿Cómo ha evolucionado el rendimiento académico de los estudiantes en el período posterior a su implementación, en comparación con las cohortes previas al SAE? ¿Qué tan grande es la diferencia que se observa entre la asignación deseada por las familias y la asignación real?

El siguiente trabajo de investigación plantea evaluar si el sistema, más allá de los principios de equidad y transparencia que inspiran su diseño teórico, está contribuyendo inadvertidamente a la segregación socioeconómica en un horizonte de mediano plazo, además, se estudia si la asignación que otorga este sistema está alejada o no de la asignación deseada por las familias.

2. Objetivos

2.1. Objetivo General

Analizar, a través de herramientas econométricas aplicadas a los datos del SIMCE y a los resultados del Sistema de Admisión Escolar, la evolución de la segregación socioeconómica, el rendimiento académico de los estudiantes y la correspondencia de los resultados con las preferencias parentales, con el fin de aportar comprensión sobre la desigualdad educativa y el papel que desempeñan en ella las políticas impulsadas por el Estado.

2.2. Objetivos Específicos

Analizar la evolución del nivel de segregación socioeconómica en Chile desde que se implementó el Sistema de Admisión Escolar, con el fin de evaluar sus efectos en la distribución de alumnos según nivel socioeconómico.

Estimar, mediante un modelo de Diferencias en Diferencias, si la evolución de los puntajes SIMCE de las cohortes post-SAE difiere de aquella observada en las cohortes previas a la implementación del sistema.

Determinar la proporción de estudiantes admitidos en el establecimiento declarado como su primera preferencia, considerando los distintos años y niveles educacionales analizados.

Comparar las distancias geográficas entre la residencia de los postulantes, el establecimiento declarado como primera preferencia y el establecimiento asignado durante la etapa regular del Sistema de Admisión Escolar.

3. Marco Teórico

3.1. Sistema Educacional en Chile

El sistema educativo chileno ha experimentado diversas transformaciones a lo largo de los años, en respuesta a los cambios económicos y sociales del país. Comprender su estructura y evolución resulta esencial para contextualizar el análisis de las políticas públicas que actualmente lo configuran. En las siguientes secciones se presentan los principales elementos que caracterizan el sistema de educación en Chile, con el propósito de entender el estudio a realizar.

3.1.1. Niveles de educación

En el contexto del sistema educativo chileno, es fundamental comprender cómo se organizan los niveles de enseñanza.

Los niveles de enseñanza comienzan por la educación parvularia, que abarca desde sala cuna hasta kínder (0 a 6 años). Su objetivo principal es favorecer el desarrollo integral de niños y niñas y promover aprendizajes pertinentes que los preparen para iniciar la educación básica. (MINEDUC, s.f.-b)

La educación básica, la cual se volvió obligatoria con la Ley de Instrucción Primaria Obligatoria, promulgada el 26 de agosto de 1920, va desde los 6 hasta los 13 años y se divide en tres niveles, el primer nivel que va desde 1° a 4° básico, el segundo nivel que va desde 5° a 6° básico y el tercer nivel que va de 7° a 8° básico. Esta etapa tiene como objetivo principal el desarrollo integral del estudiante, abarcando aspectos físicos, afectivos, sociales y cognitivos, preparándolo para la educación media.

La educación media, que va desde 1° a 4° medio, se divide en dos modalidades dependiendo

del tipo de institución donde se curse: Científico-Humanista, que está orientada en preparar a los estudiantes para la educación superior, y la Técnico-Profesional, que ofrece especialización en áreas técnicas específicas y está orientada en preparar al estudiante para el mercado laboral o para continuar sus estudios en educación superior. (Decreto N°257, 2009)

3.1.2. Tipos de financiamiento

Según lo establecido por el Ministerio de Educación (MINEDUC, s.f.-a), existen tres formas distintas de financiamiento que determinan el tipo de establecimiento educacional.

En primer lugar, se tienen los establecimientos gratuitos, los cuales dependen principalmente del financiamiento estatal y aseguran la gratuidad para los estudiantes que estudien en estos. Estos pueden ser municipales o dependientes de los servicios locales, según el tipo de administración.

Existen también los establecimientos subvencionados de financiamiento compartido, que se financian a través de la subvención estatal y a su vez con el aporte que realizan los padres mediante el pago de mensualidad, siendo la mayoría de estos establecimientos particulares.

Por otro lado, existen los establecimientos de administración delegada, los cuales operan mediante convenios de administración entre el Estado y entidades privadas vinculadas principalmente al mundo empresarial. Estos establecimientos suelen estar orientados a la formación técnico-profesional, facilitando la inserción al mercado laboral a estudiantes de sectores vulnerables.

Finalmente, los establecimientos particulares pagados son aquellos que se financian únicamente por las familias mediante el pago de mensualidades. No reciben subvención estatal y ofrecen una educación con recursos de alta calidad, siendo la opción preferida de familias de altos ingresos.

A comienzos de 1980, se emprendió en Chile una profunda reforma en el sistema educativo.

Pasando de un sistema centralizado a uno descentralizado y transfiriendo la administración de los distritos escolares a los municipios, adquiriendo un sistema de subvención por alumno, conocido como el modelo de vouchers. El Estado comenzó a entregar un subsidio fijo mensual por cada estudiante matriculado, tanto a escuelas públicas como privadas (excluyendo las privadas pagadas), siempre que no cobraran matrícula adicional (Bravo et al., 2010).

La implementación de este modelo permitió la libre entrada de nuevos colegios privados, ingresando más de 1.000 establecimientos privados al sistema, lo que generó un rápido aumento en la oferta y matrícula en escuelas privadas subvencionadas de 20 % a 40 % en 1998, disminuyendo la matrícula en escuelas públicas (Hsieh & Urquiola, 2006).

En 1993, para complementar el financiamiento estatal y diversificar la oferta, se implementó el copago, el cual consistía en cobrar una cuota adicional a los padres, además del voucher recibido por estudiante. Este permitía mejorar la infraestructura, materiales y calidad educativa, en respuesta a la presión de algunas familias que buscaban mayor inversión y diferenciación en la educación Mizala y Torche (2012).

Entre 2004 y 2010, cerca del 40 % de los colegios privados subvencionados operaban bajo copago, lo cual contribuyó a la estratificación social, concentrando a estudiantes de mayores ingresos (Gutiérrez & Carrasco, 2021). Esto llevó a que, en 2015, con la Ley de Inclusión Escolar, se buscara eliminar gradualmente la opción de que los establecimientos cobren a sus estudiantes un copago, dado que este constituía una limitante para aquellas familias que no podían costearlo (Libertad y Desarrollo, 2024).

Debido a los problemas de gestión, financiamiento y calidad asociados a la administración municipal, en 2018 se comenzó el traspaso de escuelas municipales a los Servicios Locales de Educación Pública (SLEP), tras la promulgación de la Ley N°21.040 en 2017, la cual contempla

la creación de 70 SLEP entre 2018 y 2025 (Ley N°21.040, 2017).

3.2. Sistema de Admisión Escolar

Alrededor del mundo, numerosos gobiernos han adoptado sistemas centralizados de admisión escolar con el fin de regular los procesos de selección y promover una asignación más justa de estudiantes entre los establecimientos educacionales.

Las principales razones para preferir un sistema centralizado frente a uno descentralizado son: (i) reducir la segregación socioeconómica y favorecer un acceso más equitativo a escuelas de mayor calidad; (ii) disminuir los costos y tiempos de postulación al evitar múltiples procedimientos paralelos; y (iii) promover la diversidad mediante reglas de prioridad para estudiantes de menor nivel socioeconómico (Abdulkadiroğlu & Sönmez, 2003).

En Chile, a comienzos de la década de 1980 se emprendió una profunda reforma en el sistema educativo. Pasando de un sistema centralizado a uno descentralizado y transfiriendo la administración de los distritos escolares a los municipios. Además, se estableció un voucher nacional por estudiante y se alentó a los padres a elegir libremente los colegios de sus hijos. Aquí surgieron los tres tipos de escuelas existentes en el país: (i) escuelas públicas (ii) escuelas privadas subvencionadas y (iii) escuelas privadas no subvencionadas.

Las críticas a este sistema surgieron por una mayor estratificación socioeconómica, donde la admisión a distintos tipos de escuela dependía principalmente de las características socioeconómicas del alumno (Hsieh & Urquiola, 2006).

De acuerdo con el informe elaborado por Godoy et al. (2014), los mecanismos de selección que tenían las escuelas privadas subvencionadas podían variar dependiendo del nivel del estudiante.

De prekínder a primero básico, se prohibía usar rendimiento académico o antecedentes socioeconómicos como criterio de admisión. Sin embargo, muchos establecimientos aplicaban requisitos de postulación que consistían en entrevistas a los padres, antecedentes familiares o pruebas no académicas que operaban como filtros indirectos, por lo que, aún sin historial académico, pesaba la caracterización del hogar y la capacidad de cumplir con la exigencia del colegio.

De primero a sexto básico, al igual que en los niveles menores, se prohibía usar rendimiento académico o antecedentes socioeconómicos como criterio de admisión. Aun así, las escuelas detallaban requisitos y procedimientos, legitimando pruebas propias, antecedentes familiares o compromisos económicos como condiciones de postulación.

Finalmente, de séptimo básico a cuarto medio, la selección estaba permitida, por lo que en estos niveles se intensificaban las prácticas de admisión selectiva mediante pruebas de ingreso, revisión de notas y conducta previa, entre otros filtros definidos por cada establecimiento. Esto era posible porque, al contar con más trayectoria escolar, las escuelas podían discriminar explícitamente por desempeño académico y ajuste al proyecto educativo.

Con el tiempo, este modelo generó críticas debido a su contribución al aumento de la estratificación socioeconómica. La combinación entre libre elección, financiamiento por asistencia y mecanismos de selección por parte de las escuelas dio origen a un sistema en que el tipo de establecimiento al que accedían los estudiantes estaba fuertemente determinado por sus características socioeconómicas.

Hubo dos intentos previos a la implementación del Sistema de Admisión Escolar para revertir esta situación, la primera fue en 2008, con la aprobación de la Ley N°20.248 de Subvención Escolar Preferencial, la cual consistió en entregar recursos del Estado para mejorar la equidad y calidad educativa de los establecimientos educacionales subvencionados; el segundo intento fue en

2015 con la Ley N°20.845 de Inclusión Escolar, la cual permitió avanzar decisivamente en terminar con la alta segregación escolar mediante tres disposiciones claves:

1. Instaurar un sistema centralizado de admisión escolar.
2. Un aumento gradual en el valor de los vouchers desde 2016 y eliminación paulatina del copago parental.
3. A partir de 2018, escuelas sin fines de lucro podrían recibir vouchers.

Este nuevo Sistema de Admisión Escolar centralizado fue diseñado para combatir el problema de la estratificación socioeconómica y la asignación de estudiantes se hace bajo un algoritmo de Asignación Diferida con múltiples reglas de desempate, eliminando las decisiones de selección por el lado de las escuelas con financiamiento estatal. Este algoritmo fue propuesto por primera vez por Gale y Shapley (1962) y modificado posteriormente por Abdulkadiroğlu y Sönmez (2003).

Este algoritmo considera como criterios:

- Postulantes con hermanos matriculados en alguna de las escuelas listadas.
- Hijos de funcionarios del establecimiento, hasta un mínimo de 15 % por nivel.
- Ex estudiantes que dejaron el establecimiento por diversas razones. Excepto expulsión.

Además, no hay restricciones en el número de escuelas a listar y la proximidad al establecimiento no juega ningún papel.

Una vez que los postulantes envían sus preferencias y reciben los resultados, cada estudiante tiene derecho a rechazar la escuela asignada; en ese caso, puede enviar una nueva lista y el algoritmo vuelve a ejecutarse para los cupos liberados o vacantes. Finalmente, si el algoritmo no

puede asignar al estudiante a alguna de las escuelas listadas, se le asigna automáticamente a la más cercana con cupo disponible a su residencia.

Dado que las escuelas privadas no participan en este algoritmo, los padres pueden optar por ellas en caso de no estar satisfechos con la asignación otorgada por el sistema.

La implementación de este sistema se hizo de forma gradual. Comenzó en 2016 en la región de Magallanes y para 2019 ya se encontraba implementado en la mayor parte del país.

Este sistema ha generado diversas críticas desde su implementación, tanto por parte de las familias como de expertos y actores sociales. Si bien busca mayor equidad y transparencia, persisten cuestionamientos sobre su funcionamiento y resultados.

Dentro de sus principales críticas corresponden a la pérdida de autonomía y control por parte de las familias para elegir el establecimiento y resultados percibidos como aleatorios o injustos (Bellei et al., 2025) debido a que muchas familias sienten que el sistema centralizado les quitó la capacidad de influir directamente en la elección del colegio y que el uso del algoritmo para asignar cupos, especialmente en colegios de alta demanda, es visto como un proceso poco transparente o incluso injusto, ya que no siempre se respeta la preferencia familiar ni el mérito académico.

Además de las críticas en cuanto a la no conformidad con la selección del establecimiento, la mayor parte de las críticas se debe a que su impacto en la segregación escolar ha sido modesto, persistiendo las desigualdades y habiendo incluso nuevas dificultades para grupos minoritarios Oyarzún et al. (2023). A continuación, se profundiza en el concepto de segregación escolar y en los factores que pueden estar incidiendo en su aparición.

3.3. Segregación Escolar

La adopción de un sistema centralizado de admisión en Chile responde, entre otras razones, a la necesidad de reducir la segregación escolar en Chile, entendida como la distribución desigual de estudiantes de distintos niveles socioeconómicos entre establecimientos. Esta desigualdad limita las oportunidades de acceso a una educación de calidad para estudiantes de menor nivel socioeconómico.

La segregación es un fenómeno multicausal. Además del copago, que puede excluir a hogares con menos recursos, influyen factores como la localización de las escuelas y los costos de transporte, las prioridades de acceso, las estrategias de admisión históricas y las preferencias parentales (Greaves, 2024).

En particular, las preferencias de las familias al ordenar sus postulaciones pueden reforzar patrones de separación por nivel socioeconómico. A continuación, se describen los principales criterios de elección que podrían contribuir a este resultado.

3.3.1. Elección de Establecimiento

Las preferencias de las familias en la elección de establecimientos, se forman en función de diversos criterios como la ubicación, el precio, la calidad y la competencia. (Chumacero et al., 2011)

En Chile, existe una concentración de ciertos grupos socioeconómicos en ciertas áreas de la ciudad, lo cual es conocido como segregación residencial y residentes de áreas aventajadas y desaventajadas están expuestos a distintos paquetes de servicios, donde estudiantes que residen en áreas aventajadas tienen a su disponibilidad establecimientos de mayor calidad educacional y

estudiantes de áreas desaventajadas no gozan de establecimientos con buena calidad educativa a su disposición. (Kutscher et al., 2023)

Estudios afirman que en Chile más del 95 % de los estudiantes asisten a una escuela en su comuna de residencia (Kutscher et al., 2023) reforzando la idea que la distancia al establecimiento juega un rol importante al momento de postular.

Es importante señalar que los criterios de elección pueden variar según el nivel socioeconómico de las familias, porque de ello dependen sus prioridades y restricciones. En hogares con menos recursos suelen pesar más los costos (copago, transporte, materiales). En cambio, familias con mayores recursos tienden a priorizar atributos como resultados académicos, valor agregado, proyecto educativo, oferta extracurricular o idiomas, y cuentan con más margen para asumir trayectos largos o costos adicionales.

Este comportamiento por parte de las familias intensifica la competencia entre los establecimientos, que orientan sus esfuerzos a fortalecer ciertos atributos para responder a las expectativas de los padres. A continuación, se describe cómo se ha configurado esta dinámica competitiva entre los distintos tipos de colegios.

3.4. Competencia entre establecimientos

La valoración de los resultados académicos por parte de muchas familias se traduce en una competencia entre establecimientos (sobre todo particulares subvencionados y privados) por demostrar “calidad”. En la práctica, esa calidad se comunica mediante indicadores públicos como el SIMCE, que ordenan a las escuelas en comparaciones y rankings. Este énfasis orienta el foco pedagógico y estrategias de comunicación para atraer matrícula, reforzando un mercado educativo

donde la reputación académica pesa de forma central.

Esta dinámica se conecta con la lógica competitiva instalada desde los años 80 con el sistema de vouchers, que expandió la oferta privada y ató el financiamiento al número de estudiantes matriculados. En ese contexto, competir mediante la calidad de la educación se volvió una herramienta clave para captar demanda y, con ello, recursos (Behrman et al., 2016).

A continuación se muestra la evolución de la matrícula en los principales tres tipos de establecimientos desde la completa implementación del SAE, con datos obtenidos de Ministerio de Educación, Centro de Estudios (2025):

Tabla 1: Matrícula por dependencia (2018 - 2025).

Dependencia	2018	2019	2020	2021	2022	2023	2024	2025
Municipal	1.227.163	1.236.988	1.171.352	1.112.581	1.116.914	1.112.784	1.034.121	790.189
Particular subvencionado	1.927.242	1.951.763	1.961.112	1.975.895	1.972.241	1.961.574	1.939.429	1.910.052
Particular pagado	325.500	332.675	324.860	317.821	334.438	338.213	340.318	345.910

Tabla 2: Variación de matrícula por dependencia (2018 - 2025).

Dependencia	2018-2019	2019-2020	2020-2021	2021-2022	2022-2023	2023-2024	2024-2025
Municipal	0.79 %	-5.60 %	-5.28 %	0.39 %	-0.37 %	-7.61 %	-30.87 %
Particular subvencionado	1.26 %	0.48 %	0.75 %	-0.19 %	-0.54 %	-1.14 %	-1.54 %
Particular pagado	2.16 %	-2.41 %	-2.21 %	4.97 %	1.12 %	0.62 %	1.62 %

Se puede observar una baja en la matrícula de establecimientos subvencionados, un aumento sostenido en la matrícula de establecimientos pagados y un aumento en establecimientos particulares subvencionados con una leve caída a partir de 2023.

En cuanto a la medición de la calidad, a continuación, se procede a explicar el funcionamiento del SIMCE.

3.4.1. El Sistema de Medición de la Calidad de la Educación (SIMCE)

El SIMCE, el cual se implementó en 1988 y reformado en 2012, constituye el principal sistema de evaluación estandarizada del país. Evalúa competencias en Lectura, Matemática y Ciencias Naturales mediante pruebas aplicadas anualmente y su objetivo es conocer los resultados educativos de los establecimientos, evaluando el logro de los contenidos y habilidades del Currículum Nacional (Unidad de Currículum y Evaluación, 2021).

Su diseño actual tiene 3 componentes clave:

1. Pruebas de aprendizaje.
2. Cuestionario a estudiantes, padres y docentes de factores asociados.
3. Indicadores de desarrollo personal y social.

En esta investigación, es especialmente relevante considerar el Cuestionario de Padres, dado que proporciona antecedentes detallados del contexto socio-familiar de los estudiantes (Unidad de Currículum y Evaluación, 2021).

3.5. Modelos Econométricos

Para analizar la segregación escolar en Chile antes y después del SAE, se utilizarán dos herramientas: un modelo para medir la segregación y un modelo para estimar el efecto de la implementación del SAE. A continuación, se describen los modelos a utilizar.

3.5.1. Medición de la Segregación

El índice de Duncan, conocido formalmente como índice de Disimilitud, es una de las medidas más utilizadas para cuantificar la segregación entre dos grupos en diferentes unidades, como escuelas o barrios. La popularidad de este modelo radica en su facilidad de cálculo, propiedades matemática claras e interpretación sencilla y su origen se remonta al estudio realizado por Duncan y Duncan (1955), quienes fueron los primeros en medir la segregación entre grupos ocupacionales en Estados Unidos utilizando datos del Censo de 1950.

Posterior a su origen, el índice de Duncan se consolidó como el estándar para medir la segregación residencial y ocupacional en estudios sociológicos clásicos estadounidenses como el realizado por Morgan (1980).

Formalmente, este índice se expresa de la siguiente forma:

$$D = \frac{1}{2} \sum_{i=1}^n \left| \frac{x_i}{X} - \frac{y_i}{Y} \right| \quad (1)$$

Donde:

- x_i : número de individuos del grupo A en la unidad i .
- y_i : número de individuos del grupo B en la unidad i .
- $X = \sum_{i=1}^n x_i$: total del grupo A en el área de estudio.
- $Y = \sum_{i=1}^n y_i$: total del grupo B en el área de estudio.
- n : número total de unidades (por ejemplo, barrios, distritos o escuelas).

El índice toma valores entre 0 y 1 y tiene la siguiente interpretación:

- $D = 0$ indica una distribución perfectamente integrada, es decir, no hay segregación.
- $D = 1$ indica una segregación absoluta, donde los grupos no coexisten.

En el contexto escolar, se utiliza para medir qué tan separados están, por ejemplo, los estudiantes de distintos niveles socioeconómicos entre las escuelas de una misma zona. Indica la proporción mínima de estudiantes que tendría que reubicarse para lograr una distribución equitativa entre los establecimientos, para visualizar esto se utiliza la representación del modelo hecha por Kutscher et al. (2023).

$$y_{ct} = \frac{1}{2} \sum_{j=1}^{n_{ct}} \left| \frac{T_{\text{Low}}^j}{T_{\text{Low}}} - \frac{T_{\text{High}}^j}{T_{\text{High}}} \right|, \quad (2)$$

Donde:

- y_{ct} : índice de Duncan del distrito escolar c en el año t .
- T_{Low}^j : número de estudiantes de NSE bajo en la escuela j del distrito c en el año t .
- T_{High}^j : número de estudiantes de NSE alto en la escuela j del distrito c en el año t .
- $T_{\text{Low}} = \sum_{j=1}^{n_{ct}} T_{\text{Low}}^j$ es el total de estudiantes de NSE bajo en el distrito c en el año t .
- $T_{\text{High}} = \sum_{j=1}^{n_{ct}} T_{\text{High}}^j$ es el total de estudiantes de NSE alto en el distrito c en el año t .
- n_{ct} : número de escuelas en el distrito c durante el año t .

En este estudio, los distritos escolares se definen en niveles comunales y las municipalidades se usan como sustitutos válidos para distritos escolares.

3.5.2. Diferencia de Diferencias

Diferencia de Diferencias (DiD) es uno de los métodos más populares para evaluar los efectos de políticas o intervenciones. Una de sus características que lo hacen popular al momento de evaluar efectos es que aborda el problema de endogeneidad que surge al existir factores que influyen en los resultados pero que no siempre son observados. Su popularización en la literatura se remonta a los años 70 y 80, donde se buscaban alternativas a los experimentos aleatorizados para evaluar políticas y programas Fredriksson y de Oliveira (2019). Algunos ejemplos incluyen estudios sobre el impacto de leyes laborales y reformas de salud, donde se buscaba comparar los cambios antes y después de la intervención entre grupos afectados y no afectados, como lo realizado por (Rothbard et al., 2023).

Para que este método tenga validez, es necesario usar la suposición de tendencias paralelas, es decir, que en ausencia de la intervención, ambos grupos habrían seguido trayectorias similares en el tiempo (Fredriksson & de Oliveira, 2019), además, no debe haber otras políticas que ocurran simultáneamente con la intervención, para asegurar que estas no influyan en los resultados.

La idea central del modelo es comparar cómo cambia el resultado en un grupo que recibe un tratamiento y otro grupo que no lo recibe. Su forma general se expresa de la siguiente forma:

$$y_{ist} = \beta_0 + \gamma D_{is} + \lambda T_{it} + \delta(D \cdot T)_{ist} + \epsilon_{ist} \quad (3)$$

Donde:

- y_{ist} : resultado de la unidad i del grupo s en el tiempo t
- D_{is} : dummy de grupo tratado (1 si pertenece al grupo tratado; 0 si es de control).

- T_{it} : dummy de periodo (1 después de la intervención; 0 antes)
- β_0 : intercepto correspondiente al nivel promedio del grupo de control antes del tratamiento.
- γ : coeficiente para la variable dummy que indica el periodo posterior al tratamiento.
- λ : coeficiente para la variable dummy del grupo de tratamiento.
- δ : efecto causal bajo el supuesto de tendencias paralelas y sin anticipación.
- ϵ_{ist} : término de error.

Este modelo de regresión permite un enfoque más cuantitativo y estadísticamente sólido, útil para incorporar variables adicionales que pueden influir en los resultados, mejorando la precisión del estimador δ .

4. Evolución de la segregación socioeconómica: Índice de Duncan

4.1. Descripción de Datos

Antes de proceder con la estimación de los modelos econométricos, es fundamental describir en detalle las bases de datos utilizadas en este estudio, así como el proceso de construcción de la muestra final. Esta sección tiene como objetivo presentar las fuentes de información empleadas, las variables consideradas y los criterios de selección y depuración aplicados, con el fin de asegurar la consistencia y validez del análisis.

4.1.1. Fuente de datos

El presente estudio utiliza información proveniente del Sistema de Medición de la Calidad de la Educación (SIMCE), evaluación estandarizada aplicada a nivel nacional que permite medir el rendimiento académico de los estudiantes en distintas áreas de aprendizaje. Además de los puntajes obtenidos en las pruebas, estas bases incorporan información relevante sobre características individuales, familiares y escolares, lo que permite analizar diferencias asociadas al contexto socioeconómico y educativo de los estudiantes.

En particular, se emplean las bases SIMCE de 4° básico correspondientes a los años 2016, 2017 y 2018, junto con las bases de 2° medio de los años 2022, 2023 y 2024. Esta selección permite realizar un seguimiento por cohortes, donde los estudiantes evaluados en 4° básico son observados nuevamente seis años después en 2° medio. De esta forma, las cohortes consideradas corresponden a 2016–2022, 2017–2023 y 2018–2024.

Adicionalmente, se incorporan las cohortes 2010–2016 y 2011–2017, construidas a partir de estudiantes evaluados en 4° básico antes de la implementación del Sistema de Admisión Escolar y observados posteriormente en 2° medio. La inclusión de estas cohortes permite analizar la evolución de la segregación socioeconómica en un período previo al SAE, entregando un punto de comparación para evaluar si las variaciones observadas después de la implementación del sistema corresponden a una tendencia preexistente o a cambios asociados al nuevo mecanismo de admisión.

Se utilizan también las estimaciones y proyecciones de población comunal elaboradas por el Instituto Nacional de Estadísticas (INE) para los años considerados en el estudio. Esta información se incorpora con el objetivo de construir indicadores regionales de segregación más representativos, ponderando los índices comunales según el peso poblacional de cada comuna dentro de su respectiva región.

La utilización de las cohortes permite analizar la evolución de la segregación socioeconómica escolar a lo largo de la trayectoria educativa, comparando la distribución de los estudiantes en enseñanza básica y media. Asimismo, esta estructura temporal resulta pertinente para estudiar posibles cambios asociados al funcionamiento del Sistema de Admisión Escolar (SAE), implementado gradualmente en el país desde 2016.

4.1.2. Tablas de datos utilizadas

Las bases de datos utilizadas se encuentran estructuradas en distintos conjuntos de información, organizados en tablas que contienen antecedentes específicos sobre estudiantes, establecimientos y características territoriales. Entre estas se incluyen resultados académicos individuales, información geográfica asociada a comuna, provincia y región, además de antecedentes provenien-

tes de los cuestionarios aplicados a estudiantes, padres y profesores.

Cabe señalar que la estructura y contenido de algunos cuestionarios presentan variaciones entre años, por lo que fue necesario revisar la documentación técnica y glosas correspondientes para identificar adecuadamente las variables comparables entre periodos. Asimismo, cada conjunto de información se almacena en archivos independientes, los cuales posteriormente fueron integrados mediante variables comunes para construir la base final utilizada en el análisis.

A continuación, se presentan las principales tablas empleadas en el estudio, junto con el número de observaciones y variables consideradas en cada una de ellas.

Tabla de Alumnos

Las tablas ALU corresponden a las bases de información individual de los estudiantes que rinden la evaluación SIMCE en cada año considerado. En el contexto de este estudio, su principal utilidad radica en la identificación de los alumnos dentro de las distintas cohortes analizadas.

En particular, estas tablas contienen la variable `idalumno`, la cual funciona como identificador único de cada estudiante. Esta variable permite distinguir a los alumnos dentro de la base y facilita el cruce con otras tablas del SIMCE. Si bien las tablas ALU también incluyen información asociada a puntajes y características de la evaluación, en este estudio se utilizan principalmente para reconocer y vincular a los estudiantes a lo largo del análisis.

	alu_2016	alu_2017	alu_2018	alu_2022	alu_2023	alu_2024
Número de variables	30	29	30	32	30	30
Número de observaciones	247.526	256.280	267.769	246.754	253.129	263.829

Tabla 3: Resumen de las tablas ALU utilizadas en el estudio

Tabla de Colegios

Las tablas RBD corresponden a las bases de información asociadas a los establecimientos

educacionales incluidos en cada proceso SIMCE. Estas tablas contienen antecedentes generales y de identificación de los colegios, permitiendo complementar la información individual de los estudiantes con características del establecimiento al que pertenecen.

En este estudio, las tablas se utilizan principalmente para obtener información territorial de los establecimientos educacionales. En particular, contienen variables asociadas a la región, provincia y comuna donde se ubica cada colegio. Esta información resulta relevante para el análisis, ya que permite identificar la distribución geográfica de los estudiantes y realizar posteriormente análisis de segregación socioeconómica a distintos niveles territoriales.

	rbd_2016	rbd_2017	rbd_2018	rbd_2022	rbd_2023	rbd_2024
Número de variables	41	38	42	40	42	42
Número de observaciones	7.507	7.444	7.414	2.976	2.991	3.000

Tabla 4: Resumen de las tablas RBD utilizadas en el estudio

Tabla de Encuesta de Padres

Las tablas CPAD corresponden a los cuestionarios aplicados a padres y apoderados de los estudiantes que participan en el proceso SIMCE. Estas bases contienen información relacionada con el contexto familiar y las características socioeconómicas del hogar, permitiendo complementar los antecedentes académicos y escolares de los alumnos.

En esta investigación, dichas tablas tienen especial relevancia debido a que proporcionan variables asociadas al nivel de ingresos del hogar y al nivel educacional de los padres, las cuales son utilizadas para aproximar la condición socioeconómica de los estudiantes. Asimismo, incorporan información adicional sobre características familiares y aspectos del entorno educativo que pueden influir en el desarrollo y desempeño escolar de los alumnos.

	cpad_2016	cpad_2017	cpad_2018	cpad_2022	cpad_2023	cpad_2024
Número de variables	108	111	130	101	127	131
Número de observaciones	195.575	204.134	214.111	164.984	175.321	178.266

Tabla 5: Resumen de las tablas CPAD utilizadas en el estudio

Las tablas descritas anteriormente son utilizadas de manera conjunta para la construcción del indicador de segregación considerado en este estudio. En particular, la información territorial obtenida desde las tablas RBD y las características socioeconómicas provenientes de las tablas CPAD permiten clasificar a los estudiantes según su condición socioeconómica y asociarlos a su respectivo establecimiento y ubicación geográfica. A partir de esta integración, se calcula el índice de Duncan a nivel comunal y regional, con el objetivo de medir el grado de segregación socioeconómica existente en los establecimientos educacionales.

4.2. Consolidación de Datos

En esta sección se describe el procedimiento realizado para consolidar, depurar y preparar las bases de datos utilizadas en la investigación. Se detallan las etapas de integración de las distintas fuentes de información empleadas, considerando las bases SIMCE y del Sistema de Admisión Escolar (SAE), así como los criterios aplicados para la selección y tratamiento de variables relevantes para el estudio.

Asimismo, se presentan los procesos de limpieza y transformación efectuados sobre los datos, incluyendo la eliminación de registros inconsistentes, el manejo de valores faltantes y la construcción de nuevas variables requeridas para el análisis. También se describe el proceso de vinculación entre las distintas tablas y cohortes consideradas, permitiendo generar una base consolidada que facilite el análisis.

4.2.1. Carga de datos

En primer lugar, se realizó la carga de las bases de datos correspondientes a cada año considerado en el estudio. Para ello, se utilizaron distintas funciones de lectura según el formato original de los archivos. En particular, las bases del año 2016 fueron importadas mediante la función `read_dta()` del paquete `haven`, dado que se encontraban disponibles en formato Stata.

La primera tabla cargada corresponde a `ALU`, la cual contiene información individual de los estudiantes. Las variables que contiene son:

- `agno`: Año en que se aplicó la evaluación SIMCE.
- `grado`: Nivel educacional evaluado.
- `idalumno`: Identificador único asignado a cada estudiante.
- `rbd`: Código único asociado al establecimiento educacional.
- `dvrbd`: Dígito verificador del establecimiento educacional.
- `cod_curso`: Código identificador del curso del estudiante.
- `gen_alu`: Género del estudiante.
- `ptje_mate4b_alu`: Puntaje obtenido por el estudiante en la prueba SIMCE de Matemática.
- `eem_mate4b_alu`: Error estándar de medición asociado al puntaje de Matemática.

Posteriormente, mediante la función `select()` del paquete `dplyr`, se seleccionaron únicamente las variables `idalumno` y `rbd`. Esta selección se realizó debido a que dichas variables

corresponden a los identificadores necesarios para distinguir a cada estudiante y asociarlo a su respectivo establecimiento educacional.

También se carga la tabla RBD, la cual contiene información asociada a las características y ubicación de los establecimientos educacionales. Para esta investigación, mediante la función `select()` del paquete `dplyr`, se seleccionaron únicamente las variables relevantes para el análisis territorial y socioeconómico:

- `rbd`: Código identificador único del establecimiento educacional.
- `cod_rural_rbd`: Indicador que identifica si el establecimiento pertenece a una zona urbana o rural.
- `nom_reg_rbd`: Nombre de la región donde se ubica el establecimiento.
- `nom_pro_rbd`: Nombre de la provincia donde se ubica el establecimiento.
- `nom_com_rbd`: Nombre de la comuna donde se encuentra el establecimiento.

Posteriormente, se carga la tabla CPAD, correspondiente al cuestionario aplicado a padres y apoderados. Esta base contiene información relacionada con el contexto familiar y características socioeconómicas del hogar de los estudiantes. Mediante la función `select()` del paquete `dplyr`, se seleccionaron únicamente las variables relevantes para el estudio:

- `idalumno`: Identificador único del estudiante, utilizado para realizar la vinculación con las demás tablas de información.
- `cpad_p10`: Variable asociada a la suma de los ingresos de todas las personas que aportan al hogar donde vive el estudiante, utilizada posteriormente para la construcción del indicador de segregación.

4.2.2. Fusión de las bases de datos

Una vez seleccionadas las variables relevantes de cada tabla, se procedió a fusionar las bases con el objetivo de construir una base consolidada denominada `simce`. Este proceso se realizó mediante uniones sucesivas utilizando claves comunes entre las tablas, lo que permitió integrar información individual, territorial y socioeconómica de los estudiantes.

En primer lugar, se fusionó la tabla de alumnos `ALU` con la tabla de establecimientos `RBD`, utilizando como clave común la variable `rbd`. Esta unión permitió asociar a cada estudiante con las características del establecimiento en el que se encontraba matriculado, incorporando información como dependencia administrativa, región, provincia y comuna.

Posteriormente, la base resultante se fusionó con la tabla `CPAD`, correspondiente al cuestionario de padres y apoderados. Para esta segunda unión se utilizó como clave la variable `idalumno`, permitiendo incorporar información del contexto familiar del estudiante, particularmente la suma de ingresos del hogar.

El procedimiento fue implementado mediante la función `inner_join()` del paquete `dplyr`. De esta forma, la base final conserva únicamente aquellas observaciones que se encuentran presentes en las tablas utilizadas, asegurando que cada registro cuente con información del estudiante, del establecimiento y del cuestionario de padres.

La cantidad de variables y observaciones para cada año una vez hecha la fusión de las bases se observa en la siguiente tabla.

	<code>simce_2016</code>	<code>simce_2017</code>	<code>simce_2018</code>	<code>simce_2022</code>	<code>simce_2023</code>	<code>simce_2024</code>
Número de variables	9	9	9	9	9	9
Número de observaciones	195.575	204.134	214.111	164.984	175.321	178.266

Tabla 6: Resumen de las bases `simce` consolidadas

Y para los años previos al Sistema de Admisión Escolar, las bases consolidadas contienen las siguientes observaciones y variables.

	simce_2010	simce_2011	simce_2016	simce_2017
Número de variables	9	9	9	9
Número de observaciones	222.187	213.162	167.169	169.824

Tabla 7: Resumen de las bases `simce` pre-SAE consolidadas

4.2.3. Limpieza de datos

A continuación, se realiza una etapa de depuración de la información con el objetivo de eliminar registros que pudieran introducir inconsistencias en el análisis. En particular, se revisó la variable `cpad_p10`, correspondiente a la suma de ingresos en el hogar.

Para ello, mediante la función `filter()` del paquete `dplyr`, se eliminaron aquellos registros que presentaban los códigos 0 y 99. Según la documentación del cuestionario, estos valores corresponden a respuestas inválidas o no informativas, donde:

- 0: Respuesta vacía o sin información registrada.
- 99: Doble marcación en la respuesta del cuestionario.

Asimismo, se eliminaron las observaciones con valores faltantes (NA) en la variable `cpad_p10`. Esta decisión se tomó debido a que dichos registros no permiten identificar correctamente los ingresos del hogar, los cuales son necesarios para clasificar socioeconómicamente a los estudiantes y calcular posteriormente los indicadores de segregación.

La eliminación de estos registros permite trabajar únicamente con observaciones que presenten información válida sobre los ingresos, reduciendo posibles sesgos o errores en la construcción posterior de las variables utilizadas en el análisis.

Adicionalmente, se excluyeron aquellos establecimientos cuyo valor en la variable `cod_depe2` corresponde a 3, categoría asociada a establecimientos particulares pagados. Esta decisión se tomó debido a que dichos establecimientos no participan del Sistema de Admisión Escolar (SAE), por lo que su inclusión podría sesgar el análisis de segregación asociado al sistema de admisión centralizado.

Concluida la etapa de fusión y limpieza de datos, se obtuvo una base consolidada `simce` para cada año considerado en el estudio. Estas bases contienen la información necesaria para el análisis posterior, integrando antecedentes del estudiante, del establecimiento y del contexto familiar. La cantidad final de variables y observaciones para cada año se presenta en la siguiente tabla.

	<code>simce_2016</code>	<code>simce_2017</code>	<code>simce_2018</code>	<code>simce_2022</code>	<code>simce_2023</code>	<code>simce_2024</code>
Número de variables	9	9	9	9	9	9
Número de observaciones	173.153	180.943	190.082	147.080	155.575	158.856

Tabla 8: Resumen de las bases `simce` luego de la limpieza

El mismo proceso de limpieza de datos se realizó para los datos de las cohortes previas al Sistema de Admisión Escolar.

	<code>simce_2010</code>	<code>simce_2011</code>	<code>simce_2016</code>	<code>simce_2017</code>
Número de variables	9	9	9	9
Número de observaciones	203.787	195.089	145.996	150.498

Tabla 9: Resumen de las bases `simce` pre-SAE consolidadas

4.2.4. Creación de variables

Para calcular el nivel de segregación socioeconómica, primero fue necesario clasificar a los estudiantes en dos grupos: alto nivel socioeconómico y bajo nivel socioeconómico. Esta clasificación permite comparar la distribución de ambos grupos entre establecimientos educacionales,

condición necesaria para la construcción del índice de Duncan.

La clasificación se realizó a partir de la variable `cpad_p10`, correspondiente al rango de ingresos mensuales del hogar reportado en el cuestionario de padres y apoderados. De acuerdo con la información disponible en las glosas, se clasificó como estudiantes de bajo nivel socioeconómico a aquellos pertenecientes a hogares cuyos ingresos se encontraban dentro del tramo exento del Impuesto a la Renta. En cambio, los estudiantes pertenecientes a hogares con ingresos superiores a dicho umbral fueron clasificados como de alto nivel socioeconómico.

Para implementar esta clasificación, se creó la variable `SES_ingreso` mediante la función `case_when()` del paquete `dplyr`. Esta nueva variable asigna el valor 2 a los estudiantes de bajo nivel socioeconómico y el valor 1 a los estudiantes de alto nivel socioeconómico, permitiendo posteriormente contabilizar la cantidad de estudiantes de cada grupo dentro de cada establecimiento.

4.3. Metodología

En esta sección se presenta la metodología utilizada para analizar la evolución de la segregación socioeconómica escolar y evaluar el efecto de la implementación del Sistema de Admisión Escolar (SAE). Para ello, se emplea el índice de Duncan como medida principal para cuantificar el nivel de segregación entre estudiantes de distinto nivel socioeconómico.

4.3.1. Definición de grupos socioeconómicos

Para calcular el índice de Duncan, fue necesario clasificar previamente a los estudiantes en dos grupos socioeconómicos: alto y bajo nivel socioeconómico. Esta clasificación se realizó a partir de la variable `cpad_p10`, correspondiente al rango de ingresos mensuales del hogar, disponible

en el cuestionario de padres y apoderados del SIMCE.

Se consideró como estudiantes de bajo nivel socioeconómico a aquellos pertenecientes a hogares cuyos ingresos se encontraban dentro del tramo exento del Impuesto a la Renta. Por otro lado, se clasificó como estudiantes de alto nivel socioeconómico a aquellos cuyos hogares presentaban ingresos superiores a dicho umbral.

4.3.2. Cálculo del Índice de Duncan

Una vez definidos los grupos socioeconómicos, se procedió al cálculo del índice de Duncan a nivel comunal. Este indicador permite medir el grado de segregación entre estudiantes de alto y bajo nivel socioeconómico, evaluando qué tan desigual es la distribución en ambos grupos entre los establecimientos educacionales de una misma comuna.

Formalmente, el índice de Duncan se define como:

$$D_c = \frac{1}{2} \sum_{j=1}^{J_c} \left| \frac{Low_{jc}}{T_{Low,c}} - \frac{High_{jc}}{T_{High,c}} \right|$$

donde D_c corresponde al índice de Duncan de la comuna c ; Low_{jc} representa la cantidad de estudiantes de bajo nivel socioeconómico en el establecimiento j de la comuna c ; $High_{jc}$ corresponde a la cantidad de estudiantes de alto nivel socioeconómico en dicho establecimiento; $T_{Low,c}$ es el total de estudiantes de bajo nivel socioeconómico en la comuna; $T_{High,c}$ es el total de estudiantes de alto nivel socioeconómico en la comuna; y J_c corresponde al número de establecimientos considerados dentro de la comuna.

En términos operacionales, primero se contabilizó la cantidad de estudiantes de alto y bajo nivel socioeconómico por establecimiento, comuna, provincia y región. Para ello, se agruparon

las observaciones según `nom_reg_rbd`, `nom_pro_rbd`, `nom_com_rbd`, `rbd` y la variable de clasificación socioeconómica. Luego, mediante `pivot_wider()`, las categorías socioeconómicas fueron transformadas en columnas separadas, denominadas `High` y `Low`. De esta forma, cada fila de la base representa un establecimiento dentro de una comuna, con el número de estudiantes pertenecientes a cada grupo socioeconómico.

Posteriormente, se calcularon los totales comunales de estudiantes de alto y bajo nivel socioeconómico. Estos totales corresponden a $T_{High,c}$ y $T_{Low,c}$, y fueron obtenidos sumando las columnas `High` y `Low` para todos los establecimientos pertenecientes a una misma comuna. Luego, estos totales fueron incorporados nuevamente a la base de establecimientos mediante una unión por región, provincia y comuna, permitiendo que cada establecimiento contara con el total comunal necesario para aplicar la fórmula del índice.

A continuación, se calculó para cada establecimiento el término:

$$\left| \frac{Low_{jc}}{T_{Low,c}} - \frac{High_{jc}}{T_{High,c}} \right|$$

Este término compara la proporción del total comunal de estudiantes de bajo nivel socioeconómico que se encuentra en el establecimiento j , con la proporción del total comunal de estudiantes de alto nivel socioeconómico presente en el mismo establecimiento. Mientras mayor sea la diferencia entre ambas proporciones, mayor es el aporte de ese establecimiento al nivel de segregación comunal.

Finalmente, el índice de Duncan comunal se obtuvo sumando los términos calculados para todos los establecimientos de la comuna y multiplicando el resultado por $1/2$. En aquellos casos donde una comuna no presentaba estudiantes en alguno de los dos grupos socioeconómicos, el

término fue definido como valor faltante, debido a que el índice requiere la presencia simultánea de ambos grupos para poder ser interpretado correctamente.

4.4. Estadísticas Descriptivas

En esta sección se presentan las estadísticas descriptivas de las variables utilizadas en el estudio, con el objetivo de obtener una visión general de la composición y distribución de los datos antes de realizar el análisis econométrico. En particular, se analiza la distribución socioeconómica de los estudiantes por región, la distribución territorial y ruralidad de la muestra, y la disponibilidad de establecimientos educacionales en cada territorio.

4.4.1. Distribución socioeconómica de los estudiantes por región

Con el objetivo de describir la composición socioeconómica de la muestra utilizada en el estudio, se calcularon los totales de estudiantes clasificados como de alto y bajo nivel socioeconómico para cada año analizado. Para esta clasificación se utilizó el ingreso mensual del hogar, siguiendo el criterio definido previamente en la Sección 4.3.1, según el cual se consideraron de bajo nivel socioeconómico aquellos estudiantes pertenecientes a hogares cuyos ingresos se encontraban dentro del tramo exento del Impuesto a la Renta, y de alto nivel socioeconómico aquellos cuyos ingresos superaban dicho umbral.

Este procedimiento se realizó inicialmente a nivel de establecimiento, comuna y provincia, permitiendo identificar la distribución de ambos grupos dentro de las distintas unidades territoriales consideradas.

Posteriormente, estos valores fueron tabulados a nivel regional, con el fin de presentar una visión más general de la composición socioeconómica de los estudiantes en el país. A continuación,

se muestra el porcentaje de estudiantes de bajo nivel socioeconómico, lo que permite observar diferencias territoriales relevantes y cambios en la composición de las cohortes analizadas.

Región	2010	2011	2016	2017
REGION DE ANTOFAGASTA	52,9 %	52,2 %	58,1 %	58,4 %
REGION DE ARICA Y PARINACOTA	72,8 %	74,8 %	77,4 %	74,9 %
REGION DE ATACAMA	66,4 %	64,8 %	71,5 %	68,9 %
REGION DE AYSÉN DEL GRAL. CARLOS IBÁÑEZ DEL CAMPO	68,9 %	66,7 %	65,5 %	60,5 %
REGION DE COQUIMBO	77,7 %	77,0 %	78,8 %	75,7 %
REGION DE LA ARAUCANÍA	86,5 %	86,7 %	86,6 %	84,1 %
REGION DE LOS LAGOS	82,5 %	84,0 %	83,2 %	80,7 %
REGION DE LOS RÍOS	83,5 %	84,6 %	85,0 %	81,8 %
REGION DE MAGALLANES Y DE LA ANTÁRTICA CHILENA	58,4 %	59,6 %	59,0 %	55,3 %
REGION DE ÑUBLE	—	—	—	—
REGION DE TARAPACA	62,7 %	64,1 %	68,8 %	66,1 %
REGION DE VALPARAISO	77,7 %	77,8 %	78,5 %	75,9 %
REGION DEL BIOBIO	81,7 %	82,2 %	83,6 %	81,1 %
REGION DEL LIBERTADOR GRAL. BERNARDO O'HIGGINS	81,5 %	82,0 %	83,7 %	81,6 %
REGION DEL MAULE	86,1 %	86,1 %	86,1 %	83,9 %
REGION METROPOLITANA DE SANTIAGO	71,5 %	72,0 %	71,9 %	69,2 %
Total nacional	75,85 %	76,32 %	77,33 %	74,71 %

Tabla 10: Porcentaje de estudiantes de bajo nivel socioeconómico (cohortes pre-SAE)

La Tabla 10 presenta el porcentaje de estudiantes clasificados como de bajo nivel socioeconómico para las cohortes previas a la implementación del Sistema de Admisión Escolar. Se puede observar que existe una importante heterogeneidad entre regiones. Las mayores proporciones de estudiantes de bajo nivel socioeconómico se concentran en las regiones de La Araucanía, Maule, Los Ríos, Los Lagos, Biobío y del Libertador General Bernardo O'Higgins, donde el porcentaje supera el 80 % durante la mayor parte del período analizado, situándose por sobre el porcentaje observado a nivel nacional. En contraste, las menores proporciones corresponden a Antofagasta y Magallanes, con valores cercanos al 50–60 %, lo que evidencia una composición socioeconómica más diversa.

Asimismo, se aprecia que las proporciones de estudiantes de bajo nivel socioeconómico presentan variaciones muy bajas entre los años, manteniéndose una distribución regional similar a

lo largo del período.

Finalmente, la Región de Ñuble no presenta información en esta tabla debido a que fue creada administrativamente en el año 2018, por lo que sus comunas formaban parte de la Región del Biobío durante el período pre-SAE.

Región	2016	2017	2018	2022	2023	2024
REGION DE ANTOFAGASTA	61,7 %	58,7 %	58,3 %	69,1 %	64,7 %	59,8 %
REGION DE ARICA Y PARINACOTA	76,3 %	73,0 %	71,0 %	78,6 %	75,1 %	71,4 %
REGION DE ATACAMA	69,6 %	66,6 %	66,7 %	74,7 %	66,6 %	65,9 %
REGION DE AYSÉN DEL GRAL. CARLOS IBÁÑEZ DEL CAMPO	68,2 %	61,7 %	59,5 %	68,2 %	61,9 %	58,6 %
REGION DE COQUIMBO	77,7 %	75,6 %	76,3 %	80,5 %	77,3 %	72,8 %
REGION DE LA ARAUCANÍA	86,0 %	84,1 %	81,2 %	85,0 %	82,2 %	79,0 %
REGION DE LOS LAGOS	82,4 %	79,3 %	77,7 %	79,3 %	75,8 %	71,6 %
REGION DE LOS RÍOS	84,5 %	82,9 %	79,4 %	83,0 %	79,5 %	76,3 %
REGION DE MAGALLANES Y DE LA ANTÁRTICA CHILENA	62,8 %	56,6 %	53,8 %	65,5 %	59,1 %	56,2 %
REGION DE ÑUBLE	—	—	79,0 %	84,2 %	79,8 %	76,0 %
REGION DE TARAPACA	69,4 %	65,7 %	65,4 %	74,7 %	72,3 %	68,7 %
REGION DE VALPARAÍSO	78,1 %	75,6 %	73,2 %	79,6 %	75,7 %	70,5 %
REGION DEL BIOBIO	82,1 %	78,7 %	76,6 %	80,5 %	77,3 %	72,5 %
REGION DEL LIBERTADOR GRAL. BERNARDO O'HIGGINS	82,5 %	79,4 %	77,7 %	82,3 %	80,2 %	74,2 %
REGION DEL MAULE	85,3 %	82,4 %	80,6 %	84,9 %	81,6 %	78,1 %
REGION METROPOLITANA DE SANTIAGO	72,0 %	68,3 %	66,1 %	73,9 %	69,8 %	64,4 %
Total nacional	76,78 %	73,64 %	71,79 %	77,99 %	74,18 %	73,64 %

Tabla 11: Porcentaje de estudiantes de bajo nivel socioeconómico (cohortes post-SAE)

Siguiendo con el análisis, la Tabla 11 muestra el porcentaje de estudiantes clasificados como de bajo nivel socioeconómico para las cohortes posteriores a la implementación del Sistema de Admisión Escolar. Al igual que en el período pre-SAE, se observa una marcada heterogeneidad entre regiones. Las mayores proporciones de estudiantes de bajo nivel socioeconómico se concentran principalmente en las regiones de La Araucanía, Maule, Los Ríos, Los Lagos, Biobío, Coquimbo, Valparaíso, Ñuble y del Libertador General Bernardo O'Higgins, las cuales presentan porcentajes superiores al nivel nacional en la mayoría de los años analizados. En contraste, regiones como Antofagasta y Magallanes presentan las menores proporciones, con valores cercanos al 55–70 %, reflejando una composición socioeconómica relativamente más diversa.

Al observar la evolución de las cohortes post-SAE, se aprecia un patrón relativamente si-

milar entre las distintas regiones. En las mediciones correspondientes a 4° básico, la proporción de estudiantes de bajo nivel socioeconómico presenta, en general, una disminución entre 2016 y 2018. Una tendencia similar se observa posteriormente en 2° medio, donde dicha proporción disminuye entre 2022 y 2024. Sin embargo, al pasar de las mediciones de 4° básico a las de 2° medio se observa un aumento en la proporción de estudiantes clasificados como de bajo nivel socioeconómico, comportamiento que se presenta de manera generalizada entre las regiones. A pesar de estas variaciones, la posición relativa de las regiones se mantiene relativamente estable, observándose de forma persistente mayores proporciones de estudiantes de bajo nivel socioeconómico en determinadas zonas del país.

Por lo tanto, las variaciones posteriores del índice de Duncan deben analizarse considerando tanto los cambios en la composición socioeconómica observada de las cohortes como posibles modificaciones en la distribución de los estudiantes entre establecimientos educacionales.

4.4.2. Distribución territorial y ruralidad

Con el objetivo de caracterizar territorialmente la muestra utilizada en el estudio, se analizaron las diferencias regionales en la distribución de estudiantes pertenecientes a establecimientos urbanos y rurales. Para ello, se utilizó la variable de ruralidad de la comuna, la cual permite identificar el tipo de ubicación territorial del establecimiento educacional, clasificándolo como urbano o rural.

Inicialmente, se contabilizó la cantidad de estudiantes pertenecientes a establecimientos de cada categoría a nivel regional. Posteriormente, se calcularon los porcentajes de estudiantes matriculados en establecimientos rurales para cada región y año analizado, permitiendo comparar la presencia relativa de ruralidad entre territorios y cohortes.

Las tablas que se presentan a continuación muestran el porcentaje de estudiantes pertenecientes a establecimientos rurales por región y año. Esta información permite observar diferencias importantes en la composición territorial del sistema educativo chileno, lo cual es relevante para comprender posibles heterogeneidades regionales en los niveles de segregación escolar analizados posteriormente mediante el índice de Duncan.

Región	2010	2011	2016	2017
REGION DE ANTOFAGASTA	0,6 %	2,2 %	1,4 %	1,4 %
REGION DE ARICA Y PARINACOTA	6,9 %	6,0 %	2,1 %	2,2 %
REGION DE ATACAMA	5,7 %	4,2 %	1,6 %	1,2 %
REGION DE AYSÉN DEL GRAL. CARLOS IBÁÑEZ DEL CAMPO	11,8 %	13,6 %	4,6 %	3,7 %
REGION DE COQUIMBO	18,8 %	13,3 %	2,2 %	2,2 %
REGION DE LA ARAUCANÍA	25,1 %	25,5 %	6,3 %	4,9 %
REGION DE LOS LAGOS	26,3 %	28,2 %	5,6 %	5,8 %
REGION DE LOS RÍOS	29,6 %	29,7 %	12,3 %	9,7 %
REGION DE MAGALLANES Y DE LA ANTÁRTICA CHILENA	2,3 %	3,4 %	0,0 %	0,0 %
REGION DE TARAPACA	4,1 %	17,7 %	10,2 %	12,3 %
REGION DE VALPARAISO	9,4 %	8,3 %	0,3 %	0,3 %
REGION DEL BÍO BÍO	15,3 %	14,8 %	3,2 %	3,3 %
REGION DEL LIBERTADOR GRAL. BERNARDO O'HIGGINS	24,0 %	25,7 %	3,2 %	3,8 %
REGION DEL MAULE	28,5 %	27,2 %	4,2 %	4,2 %
REGION METROPOLITANA DE SANTIAGO	5,3 %	4,1 %	2,4 %	2,2 %
Total nacional	12,9 %	12,6 %	3,2 %	3,1 %

Tabla 12: Porcentaje de estudiantes matriculados en establecimientos rurales por región y año (cohortes pre-SAE)

Se puede observar que, para las cohortes previas a la implementación del SAE, existe una importante heterogeneidad territorial en la proporción de estudiantes matriculados en establecimientos rurales. Las mayores proporciones se concentran principalmente en las regiones de Los Ríos, Maule, Los Lagos, La Araucanía y O'Higgins, especialmente durante los años 2010 y 2011, correspondientes a las mediciones de 4° básico. En estos años, dichas regiones presentan porcentajes superiores al valor nacional, que alcanza un 12,9 % en 2010 y un 12,6 % en 2011.

Al pasar a las mediciones de 2° medio, correspondientes a 2016 y 2017, la proporción de estudiantes matriculados en establecimientos rurales disminuye de manera importante en la

mayoría de las regiones. A nivel nacional, esta proporción se reduce a 3,2 % en 2016 y 3,1 % en 2017. No obstante, algunas regiones continúan presentando valores superiores al nivel nacional, destacando Los Ríos, Tarapacá, Los Lagos, La Araucanía, Maule y O'Higgins.

En particular, la Región de Magallanes presenta un porcentaje de ruralidad igual a 0 % en 2016 y 2017 dentro de la muestra analizada. Este resultado indica que, para dichos años, no se observan estudiantes de la muestra matriculados en establecimientos clasificados como rurales.

Región	2016	2017	2018	2022	2023	2024
REGION DE ANTOFAGASTA	1,8 %	2,0 %	2,0 %	0,4 %	0,2 %	0,4 %
REGION DE ARICA Y PARINACOTA	6,6 %	7,3 %	7,3 %	2,8 %	3,3 %	4,1 %
REGION DE ATACAMA	3,5 %	3,3 %	3,5 %	1,0 %	1,7 %	1,5 %
REGION DE AYSEN DEL GRAL. CARLOS IBANEZ DEL CAMPO	10,6 %	9,7 %	11,9 %	10,9 %	12,7 %	14,8 %
REGION DE COQUIMBO	12,0 %	12,0 %	12,2 %	2,8 %	2,3 %	2,4 %
REGION DE LA ARAUCANIA	21,4 %	22,0 %	21,7 %	4,0 %	4,5 %	4,2 %
REGION DE LOS LAGOS	22,9 %	23,3 %	23,9 %	7,9 %	7,5 %	7,7 %
REGION DE LOS RIOS	25,6 %	25,9 %	26,0 %	13,6 %	7,5 %	10,5 %
REGION DE MAGALLANES Y DE LA ANTÁRTICA CHILENA	2,3 %	1,6 %	1,0 %	0,0 %	0,0 %	0,0 %
REGION DE TARAPACA	17,1 %	15,5 %	16,7 %	15,8 %	16,8 %	15,1 %
REGION DE VALPARAISO	5,2 %	5,5 %	5,3 %	0,5 %	0,6 %	0,5 %
REGION DE ÑUBLE	—	—	20,4 %	7,0 %	7,2 %	6,6 %
REGION DEL BIOBIO	14,5 %	13,5 %	11,5 %	3,2 %	3,7 %	3,5 %
REGION DEL LIBERTADOR GRAL. BERNARDO O'HIGGINS	23,8 %	23,4 %	23,9 %	4,1 %	4,4 %	4,1 %
REGION DEL MAULE	26,1 %	26,1 %	26,5 %	4,4 %	4,9 %	5,2 %
REGION METROPOLITANA DE SANTIAGO	4,2 %	4,4 %	4,5 %	1,7 %	1,8 %	1,7 %
Total nacional	11,4 %	11,4 %	11,5 %	3,4 %	3,4 %	3,4 %

Tabla 13: Porcentaje de estudiantes matriculados en establecimientos rurales por región y año (cohortes post-SAE)

A partir de la Tabla 13, se observa una marcada heterogeneidad territorial en la proporción de estudiantes matriculados en establecimientos rurales. En los años 2016–2018, correspondientes a las mediciones de 4° básico, las mayores proporciones de ruralidad se concentran principalmente en las regiones del Maule, Los Ríos, O'Higgins, Los Lagos y La Araucanía. En estas regiones, los porcentajes fluctúan aproximadamente entre 21 % y 27 %, ubicándose claramente por sobre el valor nacional, que se mantiene cercano al 11,4 % durante este período. En contraste, regiones como Antofagasta, Magallanes, Atacama y la Región Metropolitana presentan proporciones considerablemente menores, lo que refleja una mayor concentración de estudiantes en establecimientos

urbanos.

En las mediciones correspondientes a 2° medio, realizadas entre 2022 y 2024, se observa una disminución importante en la proporción de estudiantes matriculados en establecimientos rurales. A nivel nacional, este porcentaje se mantiene cercano al 3,4 % durante los tres años analizados. Sin embargo, persisten diferencias territoriales relevantes. Las mayores proporciones se observan en Tarapacá, Aysén, Los Ríos y Los Lagos, mientras que regiones como Antofagasta, Magallanes, Valparaíso y la Región Metropolitana presentan valores considerablemente menores.

En términos generales, tanto en las cohortes previas como posteriores a la implementación del SAE, se observa una menor proporción de estudiantes matriculados en establecimientos rurales en 2° medio en comparación con 4° básico. Esta diferencia debe interpretarse considerando el cambio de nivel educativo entre ambos momentos, por lo que no corresponde necesariamente a una disminución temporal de la ruralidad. Una posible explicación es la menor disponibilidad de establecimientos que impartan educación media en zonas rurales, lo que podría reducir las alternativas de escolarización en dichos territorios y favorecer el desplazamiento de estudiantes hacia establecimientos urbanos.

4.4.3. Disponibilidad regional de establecimientos educacionales

Para complementar la caracterización territorial de la muestra, se analizó la cantidad de establecimientos educacionales presentes en cada región y año considerado en el estudio. Para ello, se contabilizó el número de establecimientos distintos mediante el identificador `rbd`, permitiendo obtener una aproximación a la disponibilidad de oferta escolar existente en cada territorio.

Esta información resulta relevante para el análisis de segregación socioeconómica, ya que regiones con menor cantidad de establecimientos pueden presentar mercados escolares más redu-

cidos y menores posibilidades de elección para las familias. A continuación, se presentan las tablas con el número de establecimientos educacionales por región para las cohortes previas y posteriores a la implementación del Sistema de Admisión Escolar.

Región	2010	2011	2016	2017
REGION DE ANTOFAGASTA	117	114	65	67
REGION DE ARICA Y PARINACOTA	62	65	31	31
REGION DE ATACAMA	106	102	34	34
REGION DE AYSÉN DEL GRAL. CARLOS IBÁÑEZ DEL CAMPO	54	53	24	25
REGION DE COQUIMBO	491	464	140	140
REGION DE LA ARAUCANÍA	931	891	165	166
REGION DE LOS LAGOS	740	713	144	145
REGION DE LOS RÍOS	364	363	79	78
REGION DE MAGALLANES Y DE LA ANTÁRTICA CHILENA	50	50	25	25
REGION DE TARAPACA	102	105	58	59
REGION DE VALPARAÍSO	713	716	322	323
REGION DEL BÍO-BÍO	1049	997	290	298
REGION DEL LIBERTADOR GRAL. BERNARDO O'HIGGINS	453	455	134	136
REGION DEL MAULE	603	582	158	158
REGION METROPOLITANA DE SANTIAGO	1572	1560	829	817

Tabla 14: Cantidad de establecimientos educacionales por región y año (cohortes pre-SAE)

En la Tabla 14, se observa que durante las cohortes previas a la implementación del SAE la disponibilidad de establecimientos educacionales varía considerablemente entre regiones. La Región Metropolitana concentra la mayor cantidad de establecimientos, seguida por Biobío, La Araucanía, Los Lagos y Valparaíso. Por el contrario, regiones como Magallanes, Aysén y Arica y Parinacota presentan una oferta escolar considerablemente menor.

Asimismo, se aprecia una reducción importante en el número de establecimientos al comparar los años 2010–2011, correspondientes a 4° básico, con los años 2016–2017, correspondientes a 2° medio. Este comportamiento sugiere que una parte importante de los establecimientos que imparten educación básica no ofrece continuidad hacia enseñanza media. Como consecuencia, los estudiantes enfrentan una menor disponibilidad de establecimientos al momento de continuar su

trayectoria escolar.

Región	2016	2017	2018	2022	2023	2024
REGION DE ANTOFAGASTA	110	111	111	68	67	67
REGION DE ARICA Y PARINACOTA	64	62	61	32	31	32
REGION DE ATACAMA	88	92	89	34	23	34
REGION DE AYSÉN DEL GRAL. CARLOS IBÁÑEZ DEL CAMPO	52	50	52	26	26	27
REGION DE COQUIMBO	358	353	345	134	135	135
REGION DE LA ARAUCANÍA	655	644	655	168	171	170
REGION DE LOS LAGOS	507	498	510	146	145	146
REGION DE LOS RÍOS	256	239	250	77	79	80
REGION DE MAGALLANES Y DE LA ANTÁRTICA CHILENA	46	45	43	25	25	25
REGION DE TARAPACA	100	96	93	55	55	55
REGION DE VALPARAÍSO	680	682	675	329	330	329
REGION DE ÑUBLE	–	–	262	93	93	95
REGION DEL BÍO-BÍO	831	835	569	225	225	225
REGION DEL LIBERTADOR GRAL. BERNARDO O'HIGGINS	400	409	394	136	136	142
REGION DEL MAULE	490	496	491	161	161	161
REGION METROPOLITANA DE SANTIAGO	1536	1521	1495	816	816	818

Tabla 15: Cantidad de establecimientos educacionales por región y año (cohortes post-SAE)

A partir de la Tabla 15, se observa que la disponibilidad de establecimientos, al igual que en los periodos previos al Sistema de Admisión Escolar, varía considerablemente entre regiones. La Región Metropolitana, Biobío, Valparaíso, La Araucanía y Los Lagos concentran la mayor cantidad de establecimientos, lo que es coherente con su mayor tamaño poblacional. En contraste, regiones como Magallanes, Aysén, Arica y Parinacota, Tarapacá y Atacama presentan una oferta escolar más reducida.

También se aprecia una disminución importante en el número de establecimientos al comparar los años de 4° básico con los de 2° medio. Esto puede explicarse porque no todos los establecimientos que imparten enseñanza básica ofrecen continuidad hacia enseñanza media, lo que reduce las alternativas disponibles para los estudiantes al avanzar en su trayectoria escolar. Esta diferencia es relevante para el análisis, ya que una menor disponibilidad de establecimientos puede limitar las opciones de elección de las familias y afectar la distribución socioeconómica de los estudiantes entre colegios.

En conjunto, los resultados muestran que las diferencias regionales en la disponibilidad de establecimientos constituyen una característica persistente del sistema escolar chileno, observándose tanto antes como después de la implementación del SAE. Por ello, esta variable resulta relevante para interpretar los niveles de segregación socioeconómica y analizar en qué medida las posibilidades de elección escolar dependen de la oferta educativa disponible en cada territorio.

4.5. Resultados

En esta sección se presentan los principales resultados obtenidos a partir del cálculo del índice de Duncan y de los análisis estadísticos realizados para evaluar la evolución de la segregación socioeconómica escolar en Chile. El análisis se centra en identificar diferencias territoriales y variaciones entre cohortes, así como en examinar si los cambios observados en los niveles de segregación resultan estadísticamente significativos.

A continuación, se presentan los resultados obtenidos a partir del cálculo del índice de Duncan para las cohortes analizadas. Este indicador permite evaluar el nivel de segregación socioeconómica escolar, considerando la distribución de estudiantes de alto y bajo nivel socioeconómico entre los establecimientos educacionales de cada territorio.

El análisis se desarrolla comparando los valores del índice entre los años correspondientes a 4° básico y 2° medio, con el objetivo de identificar si existen variaciones relevantes en los niveles de segregación a lo largo de la trayectoria escolar. En particular, se examinan las diferencias regionales y la evolución del indicador tanto para las cohortes previas a la implementación del SAE, correspondientes a 2010–2016 y 2011–2017, como para las cohortes posteriores, correspondientes a 2016–2022, 2017–2023 y 2018–2024.

A continuación, se presenta la tabla con los resultados obtenidos.

Región	2010	2016	Δ 2010–2016	2011	2017	Δ 2011–2017
Tarapacá	0,429	0,345	-0,084	0,390	0,325	-0,065
Antofagasta	0,373	0,370	-0,003	0,399	0,378	-0,021
Atacama	0,412	0,329	-0,083	0,388	0,325	-0,063
Coquimbo	0,490	0,389	-0,101	0,486	0,401	-0,085
Valparaíso	0,474	0,386	-0,088	0,480	0,350	-0,130
O'Higgins	0,450	0,345	-0,105	0,459	0,313	-0,146
Maule	0,518	0,396	-0,122	0,488	0,338	-0,150
Biobío	0,491	0,385	-0,106	0,499	0,377	-0,122
Araucanía	0,520	0,395	-0,125	0,531	0,372	-0,159
Los Lagos	0,517	0,413	-0,104	0,501	0,355	-0,146
Aysén	0,414	0,347	-0,067	0,461	0,311	-0,150
Magallanes	0,427	0,426	-0,001	0,400	0,368	-0,032
Metropolitana	0,420	0,364	-0,056	0,434	0,335	-0,099
Los Ríos	0,464	0,395	-0,069	0,527	0,361	-0,166
Arica y Parinacota	0,420	0,349	-0,071	0,412	0,343	-0,069

Tabla 16: Índice de Duncan por región y variación entre 4° básico y 2° medio (cohortes pre-SAE)

A partir de la Tabla 16, se observa que el índice de Duncan presenta diferencias importantes entre las regiones para los años considerados en las cohortes previas a la implementación del Sistema de Admisión Escolar. En términos generales, los valores del indicador se sitúan entre 0,31 y 0,53, lo que refleja distintos niveles de segregación socioeconómica entre los establecimientos educacionales según el territorio analizado.

Las regiones de La Araucanía, Maule, Los Lagos, Los Ríos y Biobío presentan, particularmente en las mediciones de 4° básico, algunos de los valores más elevados del índice. En contraste, regiones como Antofagasta, Tarapacá y Atacama exhiben, en general, valores relativamente menores. Estas diferencias evidencian que la segregación socioeconómica no se distribuye de manera homogénea a nivel nacional, sino que presenta una importante heterogeneidad territorial.

Las columnas de variación permiten observar, además, la evolución del índice dentro de

cada cohorte. Para la cohorte 2010–2016, todas las regiones presentan una disminución del índice de Duncan, aunque con magnitudes diferentes. Las mayores reducciones se observan en La Araucanía, Maule, Biobío, O’Higgins, Los Lagos y Coquimbo. En contraste, Antofagasta y Magallanes prácticamente no presentan variaciones.

Para la cohorte 2011–2017 se mantiene el mismo patrón general, observándose nuevamente una disminución del índice en todas las regiones. En este caso, las mayores reducciones corresponden a Los Ríos, La Araucanía, Maule, Aysén, O’Higgins y Los Lagos.

De esta forma, los resultados muestran que la disminución de la segregación entre 4° básico y 2° medio constituye un patrón común en ambas cohortes pre-SAE, aunque su magnitud varía entre regiones. Asimismo, al comparar ambas cohortes, se observa que esta disminución tiende a ser más pronunciada en la cohorte 2011–2017. En promedio, la variación regional del índice alcanza aproximadamente $-0,079$ para la cohorte 2010–2016 y $-0,102$ para la cohorte 2011–2017, evidenciando una reducción de mayor magnitud en esta última.

Con el propósito de facilitar la comparación entre cohortes y visualizar la evolución del indicador, a continuación se presentan los resultados mediante gráficos, los cuales permiten identificar con mayor claridad los cambios observados en el índice de Duncan entre 4° básico y 2° medio.

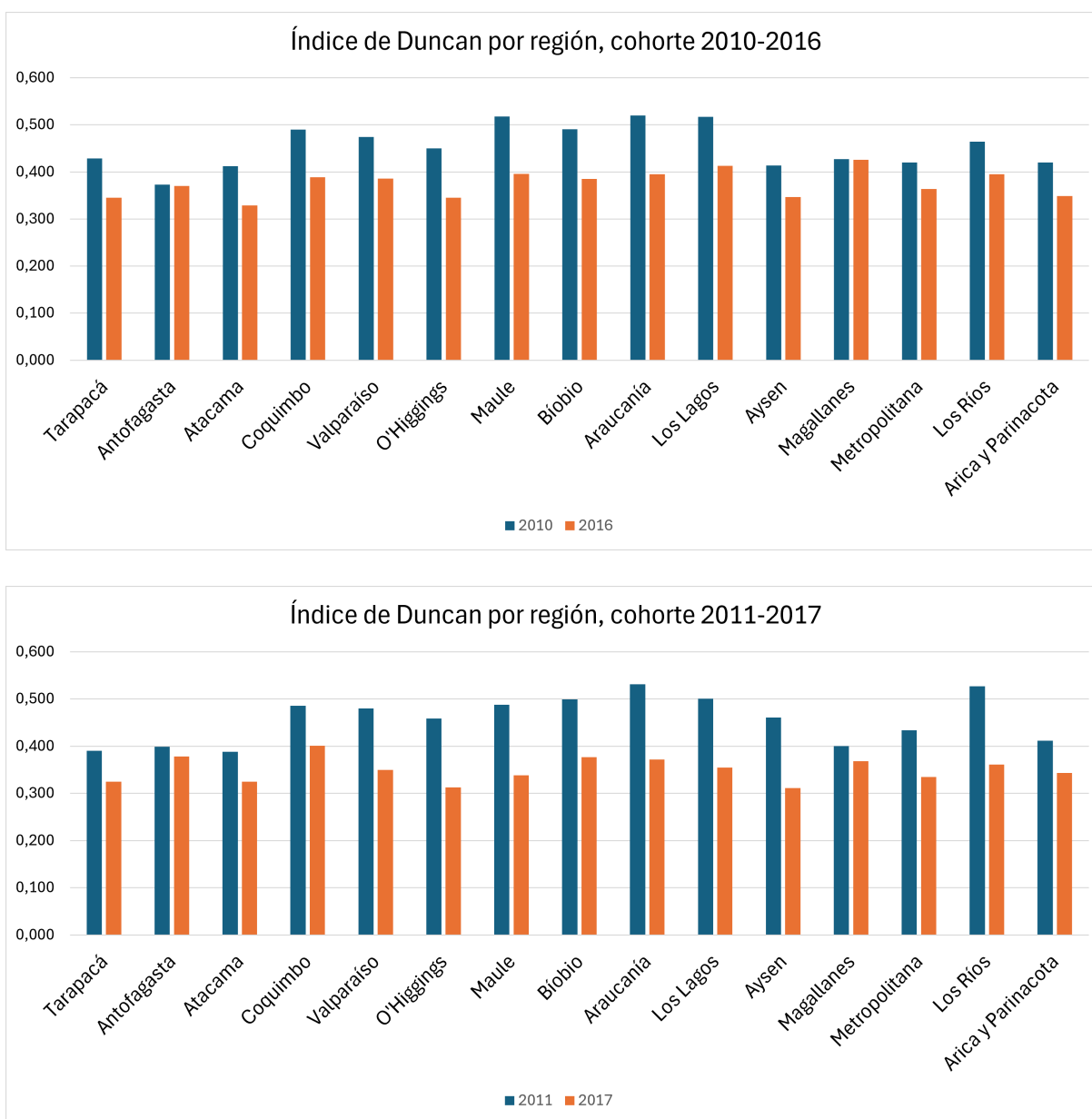


Figura 1: Índice de Duncan por región según cohorte

Los gráficos permiten comparar la evolución del índice de Duncan entre 4° básico y 2° medio para las cohortes previas a la implementación del Sistema de Admisión Escolar. En ambas cohortes (2010–2016 y 2011–2017) se observa una disminución del índice en todas las regiones, lo que indica una reducción en los niveles de segregación socioeconómica a medida que los estudiantes avanzan desde enseñanza básica hacia enseñanza media.

Si bien la magnitud de esta disminución varía entre regiones, el comportamiento general resulta consistente en ambas cohortes. Regiones como O'Higgins, Maule, Biobío, Araucanía y Los Lagos presentan reducciones relativamente mayores del índice, mientras que en regiones como Antofagasta, Magallanes y Arica y Parinacota las disminuciones entre ambos niveles son mucho menores.

En términos generales, los resultados muestran que la disminución de la segregación entre enseñanza básica y enseñanza media ya era un fenómeno observable antes de la implementación del Sistema de Admisión Escolar. Esto sugiere que parte de la reducción del índice responde a dinámicas propias de la transición entre niveles educativos, por lo que resulta necesario comparar posteriormente estos resultados con las cohortes posteriores al SAE para evaluar si la magnitud del cambio se modifica tras la implementación de esta política pública.

Región	2016	2022	Δ 2016–2022	2017	2023	Δ 2017–2023	2018	2024	Δ 2018–2024
Tarapacá	0,380	0,319	-0,061	0,351	0,324	-0,027	0,382	0,289	-0,093
Antofagasta	0,415	0,377	-0,038	0,409	0,351	-0,058	0,377	0,348	-0,029
Atacama	0,427	0,283	-0,144	0,382	0,259	-0,123	0,362	0,294	-0,068
Coquimbo	0,474	0,357	-0,117	0,421	0,325	-0,096	0,402	0,317	-0,085
Valparaíso	0,456	0,326	-0,130	0,420	0,314	-0,106	0,389	0,297	-0,092
O'Higgins	0,458	0,298	-0,160	0,395	0,281	-0,114	0,388	0,273	-0,115
Maule	0,498	0,375	-0,123	0,444	0,325	-0,119	0,405	0,293	-0,112
Biobío	0,470	0,383	-0,087	0,437	0,342	-0,095	0,418	0,325	-0,093
Araucanía	0,481	0,354	-0,127	0,461	0,325	-0,136	0,430	0,320	-0,110
Los Lagos	0,489	0,336	-0,153	0,427	0,318	-0,109	0,425	0,294	-0,131
Aysén	0,383	0,345	-0,038	0,349	0,305	-0,044	0,365	0,274	-0,091
Magallanes	0,386	0,315	-0,071	0,338	0,307	-0,031	0,318	0,286	-0,032
Metropolitana	0,400	0,309	-0,091	0,360	0,297	-0,063	0,334	0,274	-0,060
Los Ríos	0,468	0,339	-0,129	0,454	0,300	-0,154	0,458	0,308	-0,150
Arica y Parinacota	0,421	0,303	-0,118	0,410	0,332	-0,078	0,378	0,304	-0,074
Ñuble	–	0,383	–	–	0,340	–	0,474	0,309	-0,165

Tabla 17: Índice de Duncan por región y año (cohortes post-SAE)

A partir de la Tabla 17, se observa que el índice de Duncan continúa presentando diferencias importantes entre las regiones para las cohortes posteriores a la implementación del Sistema de Admisión Escolar. En términos generales, los valores del indicador se sitúan aproximadamente

entre 0,26 y 0,50, evidenciando distintos niveles de segregación socioeconómica entre los establecimientos educacionales según el territorio analizado.

Las regiones de Maule, Araucanía, Los Lagos, Los Ríos, Biobío y Coquimbo presentan, particularmente en las mediciones correspondientes a 4° básico, algunos de los valores más elevados del índice. En contraste, regiones como Tarapacá, Magallanes y la Región Metropolitana presentan valores relativamente menores. De esta forma, al igual que en las cohortes pre-SAE, se observa una marcada heterogeneidad territorial en los niveles de segregación socioeconómica.

Las columnas de variación permiten analizar la evolución del índice dentro de cada cohorte entre 4° básico y 2° medio. Para la cohorte 2016–2022, todas las regiones presentan una disminución del índice de Duncan. Las reducciones más pronunciadas se observan en O'Higgins, Los Lagos, Atacama, Valparaíso y Los Ríos. En contraste, Antofagasta y Aysén presentan las menores variaciones.

Para la cohorte 2017–2023 se mantiene el mismo patrón general, observándose nuevamente una disminución del índice en todas las regiones. En esta cohorte, las mayores reducciones corresponden a Los Ríos, La Araucanía, Atacama y Maule, mientras que Tarapacá y Magallanes presentan las disminuciones de menor magnitud.

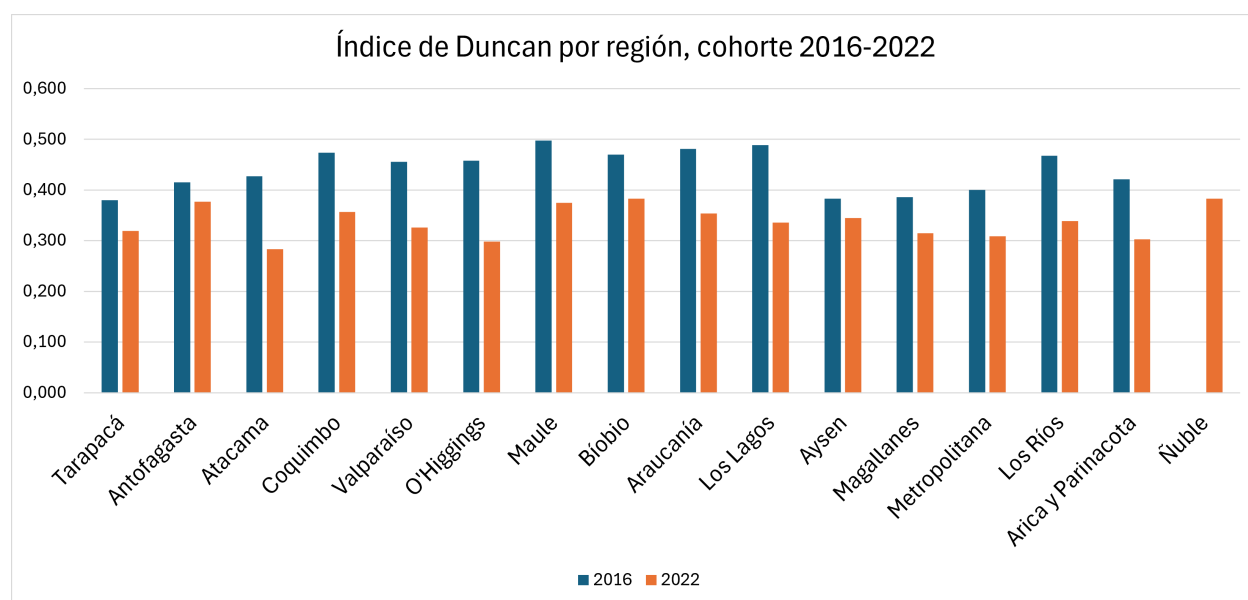
Finalmente, para la cohorte 2018–2024, también se observa una reducción generalizada de la segregación socioeconómica. La mayor disminución corresponde a Ñuble, seguida por Los Ríos, Los Lagos y O'Higgins. Por el contrario, Antofagasta y Magallanes presentan nuevamente variaciones de menor magnitud.

De esta forma, los resultados muestran que en las tres cohortes post-SAE, se observa un patrón común de disminución del índice de Duncan entre 4° básico y 2° medio, aunque su magnitud varía entre regiones y cohortes. En promedio, la variación regional alcanza aproximadamente

−0,106 para la cohorte 2016–2022, −0,090 para la cohorte 2017–2023 y −0,094 para la cohorte 2018–2024. Por lo tanto, la reducción promedio de la segregación es algo más pronunciada en la primera cohorte post-SAE, mientras que las dos cohortes posteriores presentan variaciones promedio de magnitud similar.

Al comparar estos resultados con los obtenidos para las cohortes pre-SAE, se aprecia que la disminución del índice entre 4° básico y 2° medio no constituye un fenómeno exclusivo de las cohortes posteriores a la implementación del SAE, ya que también se observa de manera generalizada en las cohortes previas. Por esta razón, la sola reducción del índice no permite atribuir directamente los cambios observados a la implementación del Sistema de Admisión Escolar, siendo necesario comparar la magnitud de las variaciones entre ambos períodos.

Con el propósito de facilitar esta comparación y visualizar con mayor claridad la evolución territorial del indicador, a continuación se presentan los resultados mediante representaciones gráficas para cada una de las cohortes post-SAE.



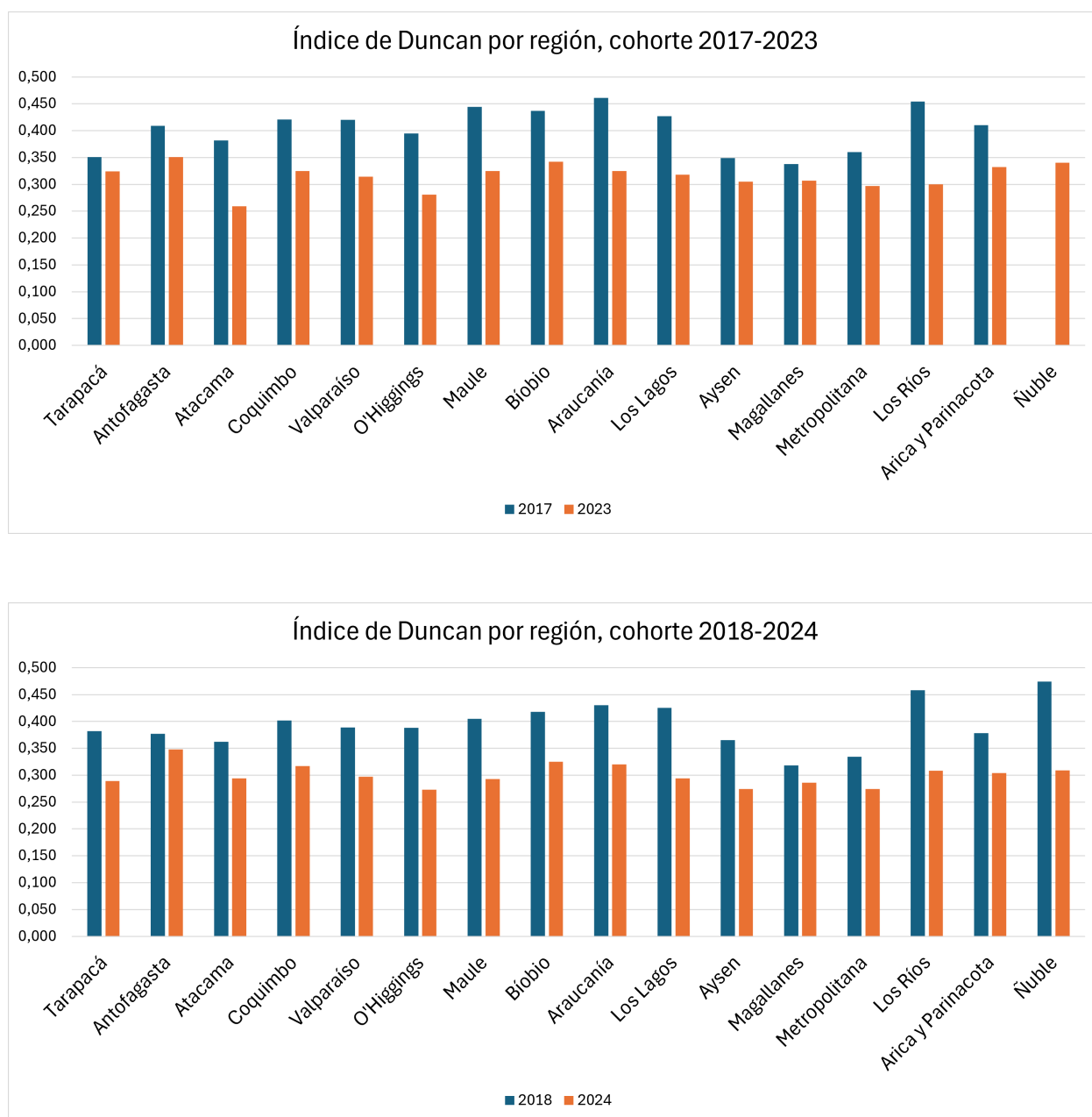


Figura 2: Índice de Duncan por región según cohorte

La Figura 2 presenta la evolución del índice de Duncan para las cohortes posteriores a la implementación del Sistema de Admisión Escolar (2016–2022, 2017–2023 y 2018–2024). Al igual que en las cohortes previas, se observa una disminución generalizada del índice al comparar los resultados de 4° básico con los de 2° medio. En particular, todas las regiones para las cuales es

posible realizar esta comparación presentan una reducción del indicador, aunque la magnitud de esta disminución varía entre regiones y cohortes.

Las reducciones más pronunciadas se observan, en general, en regiones como Los Ríos, Los Lagos, O'Higgins, La Araucanía y Maule, mientras que regiones como Antofagasta y Magallanes presentan disminuciones de menor magnitud en varias de las cohortes analizadas. Asimismo, si bien las tres cohortes muestran el mismo patrón general de reducción, la magnitud promedio de las variaciones presenta algunas diferencias, siendo la disminución algo mayor en la cohorte 2016–2022 que en las cohortes 2017–2023 y 2018–2024.

Cabe señalar que para la Región de Ñuble no se presentan valores correspondientes a 4° básico en los años 2016 y 2017, debido a que dicha región aún no había sido creada administrativamente y sus comunas formaban parte de la Región del Biobío.

En conjunto, los resultados descriptivos muestran que la disminución de la segregación socioeconómica entre 4° básico y 2° medio constituye un patrón común a las tres cohortes post-SAE. Sin embargo, este comportamiento también fue observado en las cohortes pre-SAE, por lo que la reducción del índice por sí sola no permite atribuir estos cambios a la implementación del Sistema de Admisión Escolar.

Para evaluar si las variaciones observadas son estadísticamente significativas y posteriormente determinar si existen diferencias entre las cohortes previas y posteriores al SAE, en la siguiente sección se analiza estadísticamente la variación del índice de Duncan.

En este caso, la hipótesis nula establece que la variación promedio del índice es igual a cero, mientras que la hipótesis alternativa plantea la existencia de cambios significativos en los niveles de segregación socioeconómica entre ambos periodos.

Cohorte	<i>n</i>	Variación promedio	<i>p</i> -value	Significativo al 5 %
2010–2016	15	-0.0790	9.68×10^{-7}	Sí
2011–2017	15	-0.1069	5.81×10^{-7}	Sí

Tabla 18: Resultados del test t para las variaciones del índice de Duncan (cohortes pre-SAE)

Cohorte	<i>n</i>	Variación promedio	<i>p</i> -value	Significativo al 5 %
2016–2022	16	-0.1058	5.77×10^{-8}	Sí
2017–2023	16	-0.0902	3.17×10^{-7}	Sí
2018–2024	16	-0.0938	4.80×10^{-8}	Sí

Tabla 19: Resultados del test t para las variaciones del índice de Duncan en las cohortes post-SAE

Las Tablas 18 y 19 presentan los resultados del test t aplicado a las variaciones del índice de Duncan para las cohortes previas y posteriores a la implementación del Sistema de Admisión Escolar. En ambos períodos, las variaciones promedio son negativas y estadísticamente significativas al 5 %, lo que indica que el índice de Duncan disminuye sistemáticamente entre 4° básico y 2° medio en las cohortes analizadas. Además, cabe señalar que, en la tabla 18, se considera un total de 15 regiones ($n = 15$), debido a que la Región de Ñuble aún no existía administrativamente durante los períodos analizados, por lo que sus comunas formaban parte de la Región del Biobío.

En las cohortes pre-SAE, la variación promedio alcanza un valor de $-0,0790$ para la cohorte 2010–2016 y de $-0,1069$ para la cohorte 2011–2017. Por su parte, en las cohortes post-SAE, las variaciones promedio corresponden a $-0,1058$ para 2016–2022, $-0,0902$ para 2017–2023 y $-0,0938$ para 2018–2024. En consecuencia, si bien existen diferencias en la magnitud de la reducción entre cohortes, no se observa descriptivamente un patrón en el cual las cohortes posteriores al SAE presenten sistemáticamente disminuciones mayores que las cohortes previas.

Asimismo, los *p*-values obtenidos son inferiores a 0,05 en todas las cohortes, proporcionando evidencia estadística suficiente para rechazar la hipótesis nula de que la variación promedio

del índice de Duncan es igual a cero. De esta forma, los resultados muestran que la reducción observada entre 4° básico y 2° medio constituye un patrón estadísticamente significativo tanto antes como después de la implementación del SAE.

En conjunto, estos resultados sugieren que la disminución de la segregación socioeconómica entre ambos niveles educativos ya se encontraba presente antes de la implementación del Sistema de Admisión Escolar y continúa observándose en las cohortes posteriores.

Región	<i>n</i>	Variación promedio	<i>p</i> -value	Significativo al 5 %
Tarapacá	2	-0,0745	0,0807	No
Antofagasta	2	-0,0120	0,4097	No
Atacama	2	-0,0730	0,0867	No
Coquimbo	2	-0,0930	0,0546	No
Valparaíso	2	-0,1090	0,1212	No
O'Higgins	2	-0,1255	0,1031	No
Maule	2	-0,1360	0,0653	No
Biobío	2	-0,1140	0,0446	Sí
Araucanía	2	-0,1420	0,0759	No
Los Lagos	2	-0,1250	0,1060	No
Aysén	2	-0,1085	0,2326	No
Magallanes	2	-0,0165	0,4801	No
Metropolitana	2	-0,0775	0,1723	No
Los Ríos	2	-0,1175	0,2492	No
Arica y Parinacota	2	-0,0700	0,0091	Sí

Tabla 20: Resultados del test t por región para las variaciones del índice de Duncan (cohortes pre-SAE)

Región	<i>n</i>	Variación promedio	<i>p</i> -value	Significativo al 5 %
Tarapacá	3	-0,0603	0,0869	No
Antofagasta	3	-0,0417	0,0398	Sí
Atacama	3	-0,1117	0,0388	Sí
Coquimbo	3	-0,0993	0,0088	Sí
Valparaíso	3	-0,1093	0,0101	Sí
O'Higgins	3	-0,1297	0,0134	Sí
Maule	3	-0,1180	0,0007	Sí
Biobío	3	-0,0917	0,0007	Sí
Araucanía	3	-0,1243	0,0037	Sí
Los Lagos	3	-0,1310	0,0093	Sí
Aysén	3	-0,0577	0,0751	No
Magallanes	3	-0,0447	0,0770	No
Metropolitana	3	-0,0713	0,0186	Sí
Los Ríos	3	-0,1443	0,0029	Sí
Arica y Parinacota	3	-0,0900	0,0235	Sí
Ñuble	1	-0,1650	—	No evaluable

Tabla 21: Resultados del test t por región para las variaciones del índice de Duncan (cohortes post-SAE)

Al comparar las Tablas 20 y 21, se observa que las variaciones promedio del índice de Duncan son negativas en todas las regiones, tanto para las cohortes previas como posteriores a la implementación del SAE. Esto indica que, en términos generales y considerando el conjunto de cohortes analizadas, la segregación socioeconómica tiende a disminuir entre 4° básico y 2° medio en todas las regiones consideradas.

Sin embargo, existen diferencias importantes en la significancia estadística de estas variaciones. En el período pre-SAE, únicamente las regiones del Biobío y Arica y Parinacota presentan una disminución estadísticamente significativa al 5 %. En contraste, durante el período post-SAE, 12 de las 15 regiones para las cuales es posible realizar el test presentan variaciones estadísticamente significativas. Las excepciones corresponden a Tarapacá, Aysén y Magallanes, cuyas variaciones

promedio, aunque negativas, no resultan estadísticamente significativas al 5 %. Esto sugiere que, después de la implementación del SAE, las disminuciones del índice de Duncan son más consistentes entre cohortes dentro de cada región.

En el caso de Ñuble, si bien se observa una variación de $-0,165$ para la cohorte 2018–2024, no es posible obtener un p -value ni evaluar estadísticamente dicha variación. Esto se debe a que Ñuble no existía administrativamente como región en los años 2016 y 2017, por lo que solo se dispone de una variación válida para esta región. Al contar con una única observación ($n = 1$), no es posible aplicar el test t , ya que este requiere más de una observación para estimar la variabilidad de las diferencias. Por esta razón, el resultado se reporta como “No evaluable”, lo que no implica ausencia de significancia, sino falta de observaciones suficientes para realizar la prueba estadística.

Aun así, estos resultados deben interpretarse con cautela. En el período pre-SAE se dispone únicamente de dos cohortes por región, mientras que en el período post-SAE se consideran tres cohortes, lo que entrega mayor información para estimar la variación promedio y realizar el test estadístico. Por lo tanto, las tablas permiten identificar una disminución generalizada del índice de Duncan y una mayor presencia de reducciones estadísticamente significativas en el período posterior al SAE, pero no permiten atribuir por sí solas este comportamiento a la implementación del Sistema de Admisión Escolar.

5. Efecto del SAE sobre el rendimiento académico: Puntajes

SIMCE

5.1. Descripción de Datos

En esta sección se presentan los datos utilizados para la estimación del modelo de Diferencias en Diferencias aplicado a los puntajes SIMCE, describiendo las cohortes, variables y criterios de selección empleados para construir la muestra final. Esta muestra responde a los requerimientos específicos del análisis del efecto del Sistema de Admisión Escolar sobre el rendimiento académico y, por tanto, difiere de la utilizada para el cálculo del índice de Duncan.

5.1.1. Fuente de datos

Para este análisis se utilizan las mismas fuentes de información descritas en la sección anterior, correspondientes principalmente a las bases SIMCE de 4° básico y 2° medio empleadas para construir las cohortes pre y post-SAE. En este caso, dichas bases se utilizan para realizar el seguimiento longitudinal de los estudiantes mediante su identificador individual y analizar la evolución de sus puntajes SIMCE entre ambos niveles educativos. De esta forma, se mantiene la misma estructura de cohortes previamente definida, pero orientada al estudio del rendimiento académico en lugar de la segregación socioeconómica.

5.1.2. Tablas de datos utilizadas

Para este análisis se utilizan las mismas bases de datos SIMCE descritas en la sección anterior. En particular, se emplean las bases de alumnos (ALU), establecimientos educacionales

(RBD) y cuestionarios de padres y apoderados (CPAD), cuyas principales características y variables se presentan en las Tablas 3, 4 y 5, respectivamente.

5.2. Consolidación de Datos

5.2.1. Fusión de las bases de datos

Para cada año analizado, se utilizó la base de alumnos (ALU) como base principal, incorporando posteriormente la información necesaria desde las bases CPAD y RBD. La unión con la base CPAD se realizó mediante el identificador individual del estudiante (*idalumno*), permitiendo incorporar la información de ingreso del hogar utilizada para construir la clasificación socioeconómica de los estudiantes. A partir de esta variable se definieron las categorías de alto y bajo nivel socioeconómico de acuerdo con los criterios establecidos previamente en la sección 4.3.1.

Por otra parte, la información proveniente de las bases RBD se incorporó utilizando el identificador del establecimiento (*rbd*). De esta forma, se agregaron a cada estudiante el código de comuna correspondiente a su establecimiento y el tipo de dependencia administrativa. Previo a esta unión, los códigos comunales fueron homologados entre años con el objetivo de mantener una identificación territorial consistente a lo largo del período analizado, considerando los cambios administrativos y de codificación existentes entre las distintas bases.

Como resultado de este procedimiento, para cada año se obtuvo una base ALU consolidada que contiene, además de los puntajes SIMCE y características individuales de los estudiantes, las variables socioeconómicas, de dependencia administrativa y de localización territorial necesarias para la construcción posterior de las cohortes y la estimación de los modelos.

5.2.2. Limpieza de datos

Una vez incorporada la información proveniente de las bases CPAD y RBD a la base general de alumnos (ALU), se realizó un proceso de limpieza y homologación de las variables utilizadas en el análisis. Este procedimiento tuvo como objetivo asegurar la comparabilidad entre años y conservar únicamente observaciones con información válida para la construcción de las cohortes y la posterior estimación de los modelos.

En primer lugar, para la construcción del nivel socioeconómico se utilizó la información de ingreso mensual del hogar proveniente de las bases CPAD. En los años 2010 y 2011, esta información presentaba una estructura distinta a la observada en los años posteriores, ya que el tramo de ingreso no se encontraba almacenado en una única variable categórica. En su lugar, cada alternativa de respuesta del cuestionario estaba representada mediante una variable indicadora distinta, de modo que cada estudiante contaba con un conjunto de variables binarias asociadas a los distintos tramos posibles de ingreso.

Para homologar esta información con la estructura utilizada en los años posteriores, se construyó una nueva variable categórica de ingreso. En el caso de 2010, las variables originales `cpad_p11_01` a `cpad_p11_15` correspondían a las quince alternativas de ingreso disponibles en el cuestionario. Así, cuando la variable asociada a un determinado tramo tomaba el valor 1, se asignaba a la nueva variable `cpad_p11` el número correspondiente a dicha categoría. Por ejemplo, si `cpad_p11_01=1`, el estudiante era clasificado en el tramo 1; si `cpad_p11_02=1`, en el tramo 2; y así sucesivamente hasta el tramo 15. Cuando ninguna de las categorías presentaba una respuesta válida, el ingreso se registraba como información faltante.

En la siguiente tabla se puede apreciar la construcción de la variable ingreso.

Variable original	Condición	Categoría asignada
cpad_p11_01	= 1	1
cpad_p11_02	= 1	2
cpad_p11_03	= 1	3
⋮	⋮	⋮
cpad_p11_14	= 1	14
cpad_p11_15	= 1	15
Ninguna categoría válida	–	NA

Nota: Cada variable original representa una alternativa de respuesta asociada a un tramo de ingreso. La categoría correspondiente se asignó cuando la variable indicadora tomó el valor 1. Cuando no se identificó una respuesta válida, el ingreso se registró como valor perdido (NA) y estas observaciones fueron eliminadas posteriormente.

Tabla 22: Reconstrucción de la variable categórica de ingreso en 2010

De esta manera, las múltiples variables indicadoras originales fueron consolidadas en una sola variable categórica de ingreso, permitiendo posteriormente aplicar un criterio común para la clasificación socioeconómica y facilitar la comparabilidad con los años siguientes. Para 2010 y 2011 se excluyeron los casos sin respuesta, mientras que para los años 2016, 2017, 2018, 2022, 2023 y 2024 se eliminaron además los códigos 0 y 99, correspondientes a respuestas inválidas o sin información. A partir de los casos restantes se construyó la variable de nivel socioeconómico, clasificando a los estudiantes en alto y bajo nivel socioeconómico según el criterio de ingreso definido previamente en la sección 4.3.1.

Respecto de la variable de sexo del estudiante, se homologaron las distintas codificaciones presentes entre años, de manera que el valor 1 correspondiera a hombres y el valor 2 a mujeres. Las categorías distintas de estos valores, así como los casos sin información, fueron transformadas en valores perdidos. Para la muestra final utilizada en las estimaciones se excluyeron aquellos estudiantes que no contaban con información válida de sexo.

En el caso de la comuna del establecimiento, no se eliminaron observaciones de manera directa, sino que se realizó una homologación de los códigos comunales entre años. Esto fue

necesario debido a diferencias de codificación presentes en algunas bases y a cambios administrativos ocurridos durante el período analizado. De esta forma, se buscó asegurar que una misma comuna mantuviera una identificación consistente a lo largo del tiempo antes de incorporar esta información a la base ALU.

De manera similar, la variable de dependencia administrativa del establecimiento fue incorporada desde las bases RBD mediante el identificador del establecimiento. En esta etapa no se aplicaron filtros adicionales sobre sus categorías, manteniéndose la información disponible para su posterior utilización como variable de control en las estimaciones.

Finalmente, para la construcción de las cohortes longitudinales se eliminaron los estudiantes sin identificador longitudinal válido (`mrun`). Adicionalmente, se verificó la unicidad de este identificador dentro de cada año y se excluyeron aquellos casos en que un mismo `mrun` se encontraba asociado a más de una observación, dado que en estos casos no era posible identificar de manera precisa a un único estudiante. Posteriormente, se vincularon las observaciones de 4° básico y 2° medio mediante este identificador. Para cada prueba, se mantuvieron únicamente los estudiantes con puntajes SIMCE válidos tanto en el período inicial como en el período final, asegurando así que cada individuo contara con una observación en ambos momentos de la cohorte.

5.2.3. Creación de variables

A partir de las variables originales contenidas en las distintas bases de datos, se construyeron y homologaron las variables necesarias para la posterior estimación del modelo de Diferencias en Diferencias.

En primer lugar, se creó la variable de nivel socioeconómico (`SES_control`) a partir del tramo de ingreso mensual del hogar reportado en el cuestionario de padres y apoderados. Los

estudiantes fueron clasificados en dos grupos, correspondientes a nivel socioeconómico bajo y alto, utilizando como umbral el tramo de ingreso exento del Impuesto a la Renta definido para cada período. Esta variable se definió de la siguiente manera.

- `SES_control = 1`; si el alumno pertenece al nivel socioeconómico alto.
- `SES_control = 2`; si el alumno pertenece al nivel socioeconómico bajo.

Para identificar los grupos y períodos de análisis utilizados en el modelo de Diferencias en Diferencias, se construyeron variables dicotómicas que toman valores 0 y 1. En primer lugar, la variable `tratamiento` toma el valor 1 para las cohortes posteriores a la implementación del SAE y el valor 0 para las cohortes previas, utilizadas como grupo de comparación. Por su parte, la variable `post` toma el valor 1 para las observaciones correspondientes al segundo momento de cada cohorte, es decir, 2° medio, y el valor 0 para el primer momento, correspondiente a 4° básico.

A partir de ambas variables se construyó la interacción `did`, definida como el producto entre `tratamiento` y `post`. Esta variable toma el valor 1 únicamente para las observaciones pertenecientes al grupo tratado en el período posterior, mientras que toma el valor 0 en los demás casos. El coeficiente asociado a esta interacción corresponde al estimador de Diferencias en Diferencias, ya que captura el cambio adicional observado en el grupo tratado respecto del cambio experimentado por el grupo de comparación.

5.2.4. Construcción de la base para el modelo DiD

Una vez consolidadas las bases correspondientes a cada año, se procedió a construir la base necesaria para la estimación del modelo de Diferencias en Diferencias. Para ello, los estudiantes

observados en 4° básico fueron vinculados con sus respectivas observaciones en 2° medio, utilizando el identificador longitudinal `mr_un`. Este procedimiento se realizó de manera independiente para cada una de las cinco cohortes consideradas en el análisis.

Posteriormente, las cohortes fueron reunidas en una única base longitudinal y reorganizadas de manera que cada estudiante contara con una observación correspondiente al período inicial, en 4° básico, y otra correspondiente al período final, en 2° medio. Esta estructura permitió identificar para cada observación el grupo de tratamiento, el período y su correspondiente interacción, variables utilizadas posteriormente en la estimación del modelo DiD.

Finalmente, se construyeron bases de análisis independientes para las pruebas de Lectura y Matemática, las cuales presentan la misma estructura y contienen, para cada estudiante, su identificador longitudinal (`mr_un`), cohorte de pertenencia, las variables asociadas al modelo de Diferencias en Diferencias (`tratamiento`, `post` y `did`), el puntaje SIMCE correspondiente a cada prueba y las variables utilizadas posteriormente como controles, entre ellas el nivel socioeconómico, la dependencia administrativa del establecimiento, el sexo del estudiante y la comuna. Asimismo, se conserva información asociada al año, grado y establecimiento correspondiente a cada observación.

5.3. Metodología

Para estimar el efecto de la implementación del Sistema de Admisión Escolar (SAE) sobre el rendimiento académico de los estudiantes, se utiliza el modelo de Diferencias en Diferencias (DiD). Este método permite comparar la evolución de los puntajes SIMCE entre cohortes expuestas al nuevo sistema de admisión y cohortes previas a su implementación.

La estrategia considera dos momentos de observación para cada estudiante: un período inicial, correspondiente a 4° básico, y un período posterior, correspondiente a 2° medio, seis años después. De esta manera, se compara el cambio experimentado por los estudiantes pertenecientes a las cohortes post-SAE con el cambio observado en las cohortes pre-SAE utilizadas como grupo de comparación.

5.3.1. Variables del modelo

A continuación, se describen las principales variables utilizadas en el modelo.

1. Variable Dependiente

- **puntaje:** corresponde al puntaje SIMCE obtenido por el estudiante en la prueba analizada.

En la base de Lectura, esta variable contiene los puntajes correspondientes a dicha prueba, mientras que en la base de Matemática contiene los puntajes obtenidos en Matemática. Esta constituye la variable dependiente de los modelos estimados.

2. Variables Principales del Modelo DiD

- **tratamiento:** variable dicotómica que identifica la pertenencia del estudiante al grupo de tratamiento. Toma el valor 1 para los estudiantes pertenecientes a las cohortes post-SAE (2016–2022, 2017–2023 y 2018–2024) y el valor 0 para aquellos pertenecientes a las cohortes pre-SAE (2010–2016 y 2011–2017), utilizadas como grupo de control.
- **post:** variable dicotómica que identifica el momento de observación del estudiante. Toma el valor 0 cuando la observación corresponde al período inicial de la cohorte, es decir, 4° básico, y el valor 1 cuando corresponde al período posterior, es decir, 2° medio.

- `did`: corresponde a la interacción entre las variables `tratamiento` y `post`. Esta variable toma el valor 1 únicamente para las observaciones correspondientes a estudiantes del grupo tratado en el período posterior, y 0 en los demás casos. Su coeficiente es el estimador principal del modelo de Diferencias en Diferencias, al representar el cambio experimentado por las cohortes post-SAE en comparación con el observado en las cohortes pre-SAE.

3. Variables de Control Socioeconómico y del Establecimiento

- `SES_control`: corresponde al nivel socioeconómico del estudiante, construido a partir del ingreso mensual del hogar reportado en el cuestionario de padres y apoderados. Para evitar utilizar información potencialmente afectada por acontecimientos posteriores al tratamiento, se utiliza como control la clasificación socioeconómica observada en el período inicial de cada cohorte, correspondiente a 4° básico.
- `depe_control`: identifica la dependencia administrativa del establecimiento educacional al que asistía el estudiante en el período inicial de la cohorte. Al igual que en el caso del nivel socioeconómico, se utiliza la dependencia observada en 4° básico como variable de control.
- `comuna`: identifica la comuna del establecimiento educacional correspondiente a cada observación. Esta variable se incorpora mediante efectos fijos de comuna en la especificación más completa del modelo, con el objetivo de controlar por características territoriales comunes a los establecimientos ubicados en una misma comuna.

4. Variable de Control del Estudiante

- `gen_alu`: identifica el sexo del estudiante. La variable fue recodificada de manera dicotómica, asignando el valor 0 a los hombres y el valor 1 a las mujeres. Esta variable se incorpora

explícitamente como variable de control en la especificación más completa del modelo.

5.3.2. Fórmula del modelo

El modelo econométrico utilizado para analizar el efecto del Sistema de Admisión Escolar sobre el rendimiento académico de los estudiantes se estima de manera independiente para las pruebas SIMCE de Lectura y Matemática. La especificación más completa del modelo puede representarse mediante la siguiente ecuación:

$$\begin{aligned} Y_{it} = & \beta_0 + \beta_1 \text{Tratamiento}_i + \beta_2 \text{Post}_t + \beta_3 \text{did}_{it} \\ & + \gamma \text{SES}_i + \sum_{k=2}^K \delta_k \text{Depe}_{ik} + \theta \text{Sexo}_i \\ & + \mu_{\text{comuna}(it)} + \varepsilon_{it} \end{aligned} \quad (4)$$

donde:

- Y_{it} corresponde al puntaje SIMCE obtenido por el estudiante i en el período t . El modelo se estima de manera independiente para las pruebas de Lectura y Matemática.
- β_0 corresponde a la constante del modelo.
- Tratamiento_i es una variable dicotómica que toma el valor 1 para los estudiantes pertenecientes a las cohortes post-SAE y 0 para aquellos pertenecientes a las cohortes pre-SAE.
- Post_t es una variable dicotómica que toma el valor 1 cuando la observación corresponde al período posterior, es decir, 2° medio, y 0 cuando corresponde al período inicial, 4° básico.
- $\text{Tratamiento}_i \times \text{Post}_t$ corresponde a la interacción entre las variables de tratamiento y período. El coeficiente β_3 constituye el estimador de Diferencias en Diferencias y representa

la diferencia en la evolución de los puntajes SIMCE de las cohortes post-SAE respecto de la evolución observada en las cohortes pre-SAE.

- SES_i corresponde al nivel socioeconómico del estudiante utilizado como variable de control. Para esta variable se considera la clasificación observada en el período inicial de cada cohorte, correspondiente a 4° básico.
- $Depe_{ik}$ representa las categorías de dependencia administrativa del establecimiento al que asistía el estudiante en el período inicial. Los parámetros δ_k corresponden a los coeficientes asociados a cada categoría de dependencia, considerando una de ellas como categoría de referencia.
- $Sexo_i$ corresponde al sexo del estudiante, incorporado como variable de control dicotómica, donde el valor 0 corresponde a hombres y el valor 1 a mujeres. El parámetro θ representa la diferencia promedio en el puntaje asociada a las mujeres respecto de los hombres, manteniendo constantes las demás variables del modelo.
- $\mu_{comuna(it)}$ representa los efectos fijos asociados a la comuna del establecimiento educacional correspondiente a cada observación.
- ε_{it} corresponde al término de error del modelo.

El parámetro de principal interés corresponde a β_3 , ya que permite estimar el cambio diferencial en los puntajes SIMCE de los estudiantes pertenecientes a las cohortes post-SAE respecto de las cohortes pre-SAE, una vez consideradas las variables de control y los efectos fijos incluidos en la especificación.

5.3.3. Descripción de los modelos

En el análisis econométrico se estimaron tres modelos de manera progresiva, incorporando en cada especificación nuevas variables con el objetivo de evaluar la estabilidad del efecto estimado del Sistema de Admisión Escolar sobre los puntajes SIMCE de Lectura y Matemática. Los modelos considerados se describen a continuación:

1. Modelo con las variables principales del DiD:

Este modelo corresponde a la especificación base e incluye únicamente las tres variables principales del enfoque de Diferencias en Diferencias: `tratamiento`, `post` y `did`. La variable `tratamiento` permite distinguir entre las cohortes post-SAE y las cohortes pre-SAE, mientras que `post` identifica si la observación corresponde al período inicial, en 4° básico, o al período posterior, en 2° medio. Finalmente, `did` corresponde a la interacción entre ambas variables y es el estimador principal del efecto diferencial asociado a la implementación del SAE.

Esta primera especificación permite observar la diferencia en la evolución de los puntajes SIMCE entre las cohortes post-SAE y pre-SAE sin incorporar variables de control adicionales.

2. Modelo con las variables principales + variables de control socioeconómico, del establecimiento y sexo del estudiante:

El segundo modelo incorpora, además de las variables principales del DiD, el nivel socioeconómico del estudiante (`SES_control`), la dependencia administrativa del establecimiento (`depe_control`) y el sexo del estudiante (`gen_alu`). Para estas variables se utiliza la

información correspondiente al período inicial de cada cohorte, es decir, aquella observada cuando el estudiante cursaba 4° básico.

La incorporación de estas variables permite controlar por diferencias iniciales asociadas a las condiciones socioeconómicas de los estudiantes, al tipo de establecimiento al que asistían y al sexo del estudiante, factores que pueden estar relacionados con el nivel de rendimiento académico.

3. Modelo con las variables principales + variables de control + efectos fijos por comuna:

El tercer modelo corresponde a la especificación más completa. Además de las variables incluidas en los modelos anteriores, incorpora efectos fijos por comuna del establecimiento (comuna).

La variable de sexo del estudiante se mantiene incorporada como variable de control, mientras que los efectos fijos por comuna permiten controlar por características territoriales comunes a los establecimientos ubicados dentro de una misma comuna que podrían estar relacionadas con los resultados académicos.

Esta especificación permite evaluar si el efecto estimado mediante el coeficiente δ_{id} se mantiene una vez consideradas, además de las características socioeconómicas, administrativas y del sexo del estudiante, las diferencias asociadas al contexto territorial.

La estimación progresiva de estos tres modelos permite analizar la estabilidad del coeficiente de Diferencias en Diferencias ante la incorporación de controles adicionales. De esta forma, es posible observar en qué medida la magnitud y significancia del efecto estimado del SAE sobre los puntajes SIMCE cambia al considerar distintas características de los estudiantes y de los establecimientos educacionales.

5.4. Estadísticas Descriptivas

En esta sección se presentan las principales estadísticas descriptivas de la muestra utilizada para el análisis del rendimiento académico. El objetivo es caracterizar de manera general el comportamiento de los puntajes SIMCE antes de presentar los resultados de las estimaciones econométricas. Para ello, se analiza la evolución de los puntajes de Lectura y Matemática entre 4° básico y 2° medio para las distintas cohortes consideradas en el estudio.

5.4.1. Puntajes SIMCE de lectura y matemática

A continuación, se presentan las principales estadísticas de los puntajes SIMCE de Lectura y Matemática para las cohortes consideradas en el análisis. La Tabla 23 muestra, para cada cohorte y prueba, el puntaje promedio observado en 4° básico y 2° medio, junto con su desviación estándar y la diferencia entre ambos promedios. Estas estadísticas se calculan utilizando las muestras finales construidas para cada prueba, por lo que consideran únicamente a los estudiantes que cuentan con un puntaje SIMCE válido en ambos momentos de observación.

Cohorte	Prueba	4° básico	2° medio	Δ Promedio	Desv. Est. 4°b	Desv. Est. 2°m
2010–2016	Lectura	281,69	256,32	-25,37	46,10	51,88
	Matemática	264,91	277,27	12,36	50,58	63,13
2011–2017	Lectura	277,12	261,16	-15,96	47,13	50,45
	Matemática	269,55	278,82	9,27	47,90	61,71
2016–2022	Lectura	273,66	249,51	-24,15	50,20	52,49
	Matemática	269,05	258,86	-10,19	46,74	61,84
2017–2023	Lectura	275,34	254,19	-21,15	51,91	51,49
	Matemática	267,29	263,81	-3,48	46,08	63,96
2018–2024	Lectura	277,30	255,61	-21,69	51,77	53,10
	Matemática	266,27	266,00	-0,27	47,12	65,48

Tabla 23: Estadísticas descriptivas de los puntajes SIMCE por cohorte y prueba

A partir de las estadísticas descriptivas se observan diferencias en la evolución de los puntajes entre las cohortes pre-SAE y post-SAE, especialmente en la prueba de Matemática. En Lectura,

todas las cohortes presentan un puntaje promedio menor en 2° medio respecto de 4° básico. Para las cohortes pre-SAE, las disminuciones alcanzan 25,37 puntos en 2010–2016 y 15,96 puntos en 2011–2017, mientras que en las cohortes post-SAE las reducciones fluctúan entre 21,15 y 24,15 puntos.

En Matemática, las cohortes pre-SAE presentan aumentos en el puntaje promedio entre ambos momentos, de 12,36 puntos para 2010–2016 y 9,27 puntos para 2011–2017. En contraste, las cohortes post-SAE muestran una evolución menos favorable: la cohorte 2016–2022 presenta una disminución de 10,19 puntos, la cohorte 2017–2023 una reducción de 3,48 puntos y la cohorte 2018–2024 se mantiene prácticamente estable, con una variación de -0,27 puntos.

Estas diferencias deben interpretarse con cautela, ya que no necesariamente responden a la implementación del SAE. Las cohortes post-SAE también estuvieron expuestas a otros eventos relevantes, especialmente la pandemia de COVID-19 y la interrupción de la presencialidad escolar, los cuales pudieron afectar la trayectoria de los puntajes SIMCE. En consecuencia, estos resultados muestran únicamente las diferencias observadas entre las cohortes y no permiten atribuir dichas variaciones exclusivamente al SAE.

También se observa una mayor dispersión de los puntajes en 2° medio, particularmente en Matemática, donde las desviaciones estándar superan los 60 puntos en todas las cohortes.

5.5. Resultados

En esta sección se presentan los resultados obtenidos a partir de la estimación de los modelos de Diferencias en Diferencias para las pruebas SIMCE de Lectura y Matemática. Los resultados se analizan de manera independiente para cada prueba, considerando las distintas especificaciones

definidas previamente y prestando especial atención al coeficiente asociado al estimador DiD.

5.5.1. Resultados para lectura

A continuación, se presentan y analizan los resultados obtenidos para la prueba de Lectura a partir de las tres especificaciones del modelo de Diferencias en Diferencias.

En particular, se analiza cómo varía el coeficiente asociado a la interacción entre las variables `tratamiento` y `post` a medida que se incorporan progresivamente variables relacionadas con las características socioeconómicas de los estudiantes, la dependencia administrativa del establecimiento y los efectos fijos por sexo y comuna. De esta forma, se busca determinar en qué medida la incorporación de estos controles modifica la relación estimada entre la implementación del SAE y la evolución de los puntajes SIMCE de Lectura.

Tabla 24: Resumen de resultados de las regresiones para Lectura

Variables	(1) Modelo 1	(2) Modelo 2	(3) Modelo 3
Tratamiento	-3,856*** (0,131)	-4,718*** (0,127)	-4,676*** (0,127)
Post	-20,570*** (0,100)	-20,570*** (0,100)	-21,377*** (0,102)
DiD	-1,693*** (0,126)	-1,693*** (0,126)	-1,552*** (0,126)
Intercepto	279,369*** (0,099)	275,493*** (0,181)	— —
Control por nivel socioeconómico	No	Sí	Sí
Control por dependencia	No	Sí	Sí
Control por sexo	No	Sí	Sí
Efectos fijos por comuna	No	No	Sí
Observaciones	1.151.674	1.151.674	1.151.674
R^2 ajustado	0,0452	0,1236	0,1341

Notas: *** $p < 0,01$, ** $p < 0,05$, * $p < 0,10$. Los errores estándar, presentados entre paréntesis, se encuentran agrupados a nivel del estudiante mediante el identificador `mrun`.

Tabla 25: Resultados de las variables de control para Lectura

Variable	(2) Modelo 2	(3) Modelo 3
Nivel socioeconómico bajo	-15,770*** (0,136)	-15,324*** (0,138)
Particular Subvencionado	11,724*** (0,128)	12,580*** (0,136)
Particular	30,638*** (0,234)	29,119*** (0,256)
Servicios Locales de Educación Pública	2,575** (1,031)	4,635*** (1,053)
Mujer	10,762*** (0,114)	10,689*** (0,113)
Efectos fijos por comuna	No	Sí

Notas: *** $p < 0,01$, ** $p < 0,05$, * $p < 0,10$. La categoría de referencia para la variable sexo corresponde a los hombres. Los errores estándar, presentados entre paréntesis, se encuentran agrupados a nivel del estudiante mediante `mrun`.

Primera regresión. En la primera especificación se consideran únicamente las variables principales del modelo de Diferencias en Diferencias: `tratamiento`, `post` y `did`. Los resultados muestran que el coeficiente asociado a `tratamiento` es negativo y estadísticamente significativo ($-3,856$; $p < 0,01$), indicando que, en el período inicial, las cohortes post-SAE presentan en promedio puntajes de Lectura aproximadamente 3,86 puntos inferiores a los observados en las cohortes pre-SAE. Por su parte, el coeficiente de `post` también es negativo y significativo ($-20,570$; $p < 0,01$), lo que refleja una disminución promedio de los puntajes entre 4° básico y 2° medio para las cohortes utilizadas como grupo de comparación.

El coeficiente de principal interés corresponde a la interacción `did`, cuyo valor es de $-1,693$ y resulta estadísticamente significativo al 1 %. Este resultado indica que la variación observada en los puntajes de Lectura de las cohortes post-SAE es aproximadamente 1,69 puntos más negativa que aquella experimentada por las cohortes pre-SAE entre ambos períodos. El R^2 ajustado del modelo es de 0,0452, lo que indica que esta especificación base explica una proporción acotada de la variabilidad total de los puntajes de Lectura.

Segunda regresión. En la segunda especificación se incorporan como variables de control el nivel socioeconómico del estudiante, la dependencia administrativa del establecimiento y el sexo del estudiante. Al incluir estos controles, el coeficiente asociado a `tratamiento` disminuye a $-4,718$ y continúa siendo estadísticamente significativo ($p < 0,01$). El coeficiente de `post` se mantiene en $-20,570$, mientras que el estimador de Diferencias en Diferencias permanece en $-1,693$ y continúa siendo significativo al 1 %. Por lo tanto, la incorporación de estas características no modifica sustancialmente el efecto estimado por el coeficiente DiD respecto de la primera especificación.

En relación con las variables de control, los estudiantes clasificados en el grupo socioeconómico bajo presentan, en promedio, puntajes aproximadamente 15,77 puntos inferiores a los estudiantes pertenecientes al grupo socioeconómico alto, manteniendo constantes las demás variables del modelo. Asimismo, se observan diferencias significativas entre las distintas categorías de dependencia administrativa respecto de la categoría utilizada como referencia, la cual corresponde a establecimientos municipales. En cuanto al sexo del estudiante, las mujeres presentan, en promedio, un puntaje de Lectura 10,76 puntos superior al de los hombres, manteniendo constantes las demás variables incluidas en el modelo. Este coeficiente es estadísticamente significativo al 1 %. La incorporación de estos controles aumenta el R^2 ajustado desde 0,0452 a 0,1236, reflejando una mayor capacidad del modelo para explicar las diferencias observadas en los puntajes de Lectura.

Tercera regresión. Finalmente, la tercera especificación incorpora, además de las variables incluidas previamente, efectos fijos por comuna. En este modelo, el coeficiente de `tratamiento` alcanza un valor de $-4,676$ y continúa siendo estadísticamente significativo ($p < 0,01$). Por su parte, el coeficiente de `post` corresponde a $-21,377$, manteniéndose negativo y significativo.

El coeficiente asociado a la interacción `did` disminuye levemente en magnitud respecto

de las especificaciones anteriores, pasando de $-1,693$ a $-1,552$, aunque continúa siendo negativo y estadísticamente significativo al 1 %. De esta forma, incluso después de controlar por nivel socioeconómico, dependencia administrativa y sexo del estudiante, además de incorporar efectos fijos por comuna, las cohortes post-SAE presentan una evolución del puntaje de Lectura aproximadamente 1,55 puntos más negativa que las cohortes pre-SAE. La estabilidad del signo, magnitud y significancia del estimador DiD entre las distintas especificaciones indica que el resultado es relativamente robusto a la incorporación progresiva de las variables de control consideradas.

En cuanto a los controles, el coeficiente asociado al nivel socioeconómico bajo se mantiene negativo y significativo, con una diferencia aproximada de 15,32 puntos respecto del grupo socioeconómico alto. Las categorías de dependencia administrativa también mantienen coeficientes estadísticamente significativos. Asimismo, el coeficiente asociado a la variable de sexo indica que las mujeres presentan, en promedio, un puntaje de Lectura aproximadamente 10,69 puntos superior al de los hombres, manteniendo constantes las demás variables del modelo y considerando los efectos fijos por comuna. Finalmente, el R^2 ajustado aumenta a 0,1341, mostrando una mejora adicional en el ajuste del modelo tras incorporar los efectos fijos por comuna.

5.5.2. Resultados para matemática

Al igual que en el análisis realizado para la prueba de Lectura, a continuación se presentan los resultados obtenidos para Matemática a partir de las tres especificaciones del modelo de Diferencias en Diferencias. En este caso, el análisis se centra en la evolución del coeficiente DiD y en su estabilidad ante la incorporación progresiva de las variables de control y efectos fijos considerados en los modelos.

Variables	(1) Modelo 1	(2) Modelo 2	(3) Modelo 3
Tratamiento	0,194 (0,130)	-1,226*** (0,124)	-1,134*** (0,123)
Post	10,777*** (0,103)	10,777*** (0,103)	9,679*** (0,104)
DiD	-15,195*** (0,134)	-15,195*** (0,134)	-14,914*** (0,134)
Intercepto	267,283*** (0,104)	271,304*** (0,191)	— —
Control por nivel socioeconómico	No	Sí	Sí
Control por dependencia	No	Sí	Sí
Control por sexo	No	Sí	Sí
Efectos fijos por comuna	No	No	Sí
Observaciones	1.167.118	1.167.118	1.167.118
R^2 ajustado	0,0086	0,1351	0,1556

Notas: *** $p < 0,01$, ** $p < 0,05$, * $p < 0,10$. Los errores estándar, presentados entre paréntesis, se encuentran agrupados a nivel del estudiante mediante el identificador `mrun`.

Tabla 26: Resultados de las variables de control para Matemática

Variable	(2) Modelo 2	(3) Modelo 3
Nivel socioeconómico bajo	-20,248*** (0,146)	-19,306*** (0,147)
Particular Subvencionado	17,263*** (0,135)	17,318*** (0,143)
Particular	49,579*** (0,254)	44,719*** (0,279)
Servicios Locales de Educación Pública	4,368*** (1,016)	7,294*** (1,033)
Mujer	-7,022*** (0,121)	-7,111*** (0,119)
Efectos fijos por comuna	No	Sí

Notas: *** $p < 0,01$, ** $p < 0,05$, * $p < 0,10$. La categoría de referencia para la variable sexo corresponde a los hombres. Los errores estándar, presentados entre paréntesis, se encuentran agrupados a nivel del estudiante mediante `mrun`.

Primera regresión. En la primera especificación se consideran únicamente las variables principales del modelo de Diferencias en Diferencias: `tratamiento`, `post` y `did`. Los resultados muestran que el coeficiente asociado a `tratamiento` es positivo, con un valor de 0,194, pero no

resulta estadísticamente significativo. Esto indica que, en el período inicial, no se observan diferencias estadísticamente significativas en los puntajes de Matemática entre las cohortes post-SAE y las cohortes pre-SAE utilizadas como grupo de comparación.

Por su parte, el coeficiente de `post` es positivo y estadísticamente significativo (10,777; $p < 0,01$), lo que indica que, para las cohortes pre-SAE, los puntajes de Matemática son en promedio aproximadamente 10,78 puntos superiores en 2° medio respecto de 4° básico.

El coeficiente de principal interés corresponde a la interacción `did`, cuyo valor es de $-15,195$ y resulta estadísticamente significativo al 1 %. Este resultado indica que la evolución de los puntajes de Matemática de las cohortes post-SAE es aproximadamente 15,20 puntos más negativa que la observada en las cohortes pre-SAE entre ambos períodos. El R^2 ajustado del modelo es de 0,0086, lo que muestra que esta especificación base explica una proporción reducida de la variabilidad total de los puntajes de Matemática.

Segunda regresión. En la segunda especificación se incorporan como variables de control el nivel socioeconómico del estudiante, la dependencia administrativa del establecimiento y el sexo del estudiante. Al incluir estos controles, el coeficiente asociado a `tratamiento` pasa a ser negativo, alcanzando un valor de $-1,226$, y resulta estadísticamente significativo al 1 %. Esto indica que, una vez consideradas estas características, las cohortes post-SAE presentan en el período inicial puntajes de Matemática aproximadamente 1,23 puntos inferiores a los de las cohortes pre-SAE.

El coeficiente de `post` se mantiene en 10,777, mientras que el estimador de Diferencias en Diferencias permanece en $-15,195$ y continúa siendo estadísticamente significativo al 1 %. De esta forma, la incorporación de las variables de control no modifica la magnitud del coeficiente DiD respecto de la primera especificación.

En relación con las variables de control, los estudiantes clasificados en el grupo socioeconómico bajo presentan, en promedio, puntajes aproximadamente 20,25 puntos inferiores a los estudiantes pertenecientes al grupo socioeconómico alto, manteniendo constantes las demás variables del modelo. Asimismo, se observan diferencias estadísticamente significativas entre las distintas categorías de dependencia administrativa respecto de la categoría de referencia, correspondiente a establecimientos municipales. En cuanto al sexo del estudiante, las mujeres presentan, en promedio, puntajes de Matemática aproximadamente 7,02 puntos inferiores a los de los hombres, manteniendo constantes las demás variables incluidas en el modelo. Este coeficiente es estadísticamente significativo al 1 %. La incorporación de estas variables incrementa el R^2 ajustado desde 0,0086 a 0,1351, reflejando una mejora importante en la capacidad explicativa del modelo.

Tercera regresión. Finalmente, la tercera especificación incorpora, además de las variables incluidas previamente, efectos fijos por comuna. En este modelo, el coeficiente asociado a `tratamiento` alcanza un valor de $-1,134$ y continúa siendo estadísticamente significativo al 1 %. Por su parte, el coeficiente de `post` disminuye levemente a $9,679$, manteniéndose positivo y significativo.

El coeficiente asociado a la interacción `did` disminuye ligeramente en magnitud, pasando de $-15,195$ a $-14,914$, aunque continúa siendo negativo y estadísticamente significativo al 1 %. Esto indica que, incluso después de controlar por nivel socioeconómico, dependencia administrativa y sexo del estudiante, además de incorporar efectos fijos por comuna, las cohortes post-SAE presentan una evolución del puntaje de Matemática aproximadamente 14,91 puntos más negativa que las cohortes pre-SAE. La estabilidad del signo, magnitud y significancia del estimador DiD entre las distintas especificaciones muestra que el resultado se mantiene relativamente robusto ante la incorporación progresiva de controles.

En cuanto a las variables de control, el coeficiente asociado al nivel socioeconómico bajo se mantiene negativo y significativo, con una diferencia aproximada de 19,31 puntos respecto del grupo socioeconómico alto. Las categorías de dependencia administrativa también mantienen coeficientes positivos y estadísticamente significativos respecto de la categoría de referencia. Finalmente, el R^2 ajustado aumenta a 0,1556, mostrando una mejora adicional en el ajuste del modelo tras incorporar los efectos fijos por comuna.

Finalmente, al comparar los resultados obtenidos para Lectura y Matemática, se observan algunas similitudes y diferencias relevantes entre ambas pruebas. En los dos casos, el coeficiente asociado a `did` es negativo y estadísticamente significativo en las tres especificaciones, manteniendo además una magnitud relativamente estable frente a la incorporación progresiva de variables de control y efectos fijos por comuna. Sin embargo, la magnitud de la diferencia estimada es considerablemente mayor en Matemática, donde el coeficiente DiD se sitúa en torno a -15 puntos, mientras que en Lectura alcanza valores cercanos a -1,5 puntos. Asimismo, mientras en Lectura el coeficiente de `post` es negativo en todas las especificaciones, en Matemática es positivo, reflejando trayectorias distintas entre ambas pruebas. En cuanto a los controles, el nivel socioeconómico bajo presenta un coeficiente negativo y estadísticamente significativo en ambos casos, aunque de mayor magnitud en Matemática. Finalmente, la incorporación de controles y efectos fijos por comuna mejora el ajuste de los modelos en ambas pruebas, siendo este incremento especialmente evidente en Matemática.

No obstante, estos resultados deben interpretarse con cautela, ya que las diferencias estimadas no necesariamente pueden atribuirse exclusivamente a la implementación del SAE. Las cohortes post-SAE estuvieron expuestas a la pandemia de COVID-19, a la interrupción prolongada de la presencialidad escolar y al posterior proceso de retorno a clases presenciales, factores que

pudieron afectar de manera importante la trayectoria de los aprendizajes. Del mismo modo, las cohortes pre-SAE también estuvieron expuestas a eventos particulares. En especial, el terremoto ocurrido en Chile en 2010 pudo haber afectado temporalmente el desempeño de algunos estudiantes en el período inicial, particularmente en las zonas más afectadas. En caso de que dicho evento haya reducido de manera transitoria los puntajes de 4° básico, la posterior recuperación en 2° medio podría reflejar parcialmente un efecto de normalización, generando un aumento observado que no necesariamente correspondería únicamente a una trayectoria académica habitual. Por lo tanto, tanto la pandemia como otros eventos ocurridos en los períodos analizados deben considerarse al interpretar la magnitud y dirección de los coeficientes estimados.

6. Asignación escolar y primera preferencia en el Sistema de Admisión Escolar

6.1. Descripción de Datos

Para analizar la relación entre las preferencias declaradas por las familias y la asignación finalmente obtenida a través del Sistema de Admisión Escolar, es necesario describir las fuentes de información utilizadas y la forma en que estas fueron vinculadas. En esta sección presentan las bases de datos consideradas.

6.1.1. Fuente de datos

El análisis utiliza información proveniente de las bases administrativas del Sistema de Admisión Escolar (SAE) correspondientes a los años posteriores a su implementación considerados en este estudio. Estas bases contienen información asociada al proceso de postulación de los estudiantes, incluyendo los establecimientos incorporados en sus preferencias, el orden en que fueron declarados y el establecimiento asignado al término del proceso de la etapa regular.

En particular, se emplean las bases del SAE correspondientes a los años 2016, 2017, 2018, 2022, 2023 y 2024. La selección de estos años busca mantener consistencia temporal con las cohortes post-SAE analizadas previamente en la investigación, permitiendo complementar los resultados de segregación socioeconómica y rendimiento académico con información sobre el proceso de asignación escolar en los mismos períodos de referencia. De esta manera, este análisis incorpora una dimensión adicional del funcionamiento del SAE, relacionada con las preferencias declaradas por las familias y los establecimientos asignados en la etapa regular.

6.1.2. Tablas de datos

Para cada año analizado se utilizaron las bases de datos disponibles del Sistema de Admisión Escolar correspondientes a la etapa regular del proceso de admisión. En particular, se emplearon las tablas A1, B1, C1 y D1, las cuales contienen información sobre la oferta de establecimientos, los postulantes, las preferencias declaradas y los resultados de asignación, respectivamente.

Estas bases corresponden a la etapa regular del proceso de postulación y asignación. Por lo tanto, no se consideran en este análisis las bases correspondientes a la etapa complementaria, en la cual pueden participar estudiantes que no postularon durante la etapa regular o que rechazaron la asignación obtenida en dicha etapa. A continuación, se describe el contenido de cada una de las bases utilizadas.

1. A1: Oferta de establecimientos

Esta base contiene información relativa a la oferta de establecimientos educacionales disponible durante la etapa regular del Sistema de Admisión Escolar. Además, esta base contiene información sobre los cupos y vacantes disponibles, así como la ubicación geográfica de los establecimientos.

Entre las principales variables disponibles en esta base se encuentran:

- RBD: identificador del establecimiento educacional.
- COD_NIVEL: identifica el nivel educativo ofrecido por el establecimiento.
- COD_ENSE y COD_GRADO: identifican el tipo de enseñanza y el grado correspondiente.

- LAT y LON: corresponden a las coordenadas geográficas de latitud y longitud del establecimiento.
- CUPOS_TOTALES: representa el total de cupos declarados por el establecimiento para una determinada combinación de establecimiento y nivel.
- VACANTES: corresponde al número de vacantes disponibles para el proceso de admisión.
- VACANTES_REGULAR: identifica las vacantes destinadas al proceso regular de admisión.

2. B1: Postulantes etapa regular

Esta base contiene información individual de los estudiantes que participan en la etapa regular del Sistema de Admisión Escolar. Esta base permite identificar a cada postulante mediante su MRUN, conocer el nivel educacional al que postula y disponer de información geográfica aproximada asociada a su lugar de residencia.

Entre las principales variables disponibles en esta base se encuentran:

- MRUN: identificador del postulante, utilizado para vincular información del estudiante entre las distintas bases del SAE.
- COD_NIVEL: identifica el nivel educacional al que postula el estudiante.
- LAT_CON_ERROR y LON_CON_ERROR: corresponden a las coordenadas geográficas aproximadas del hogar del postulante, incorporando un error aleatorio con el fin de resguardar la privacidad de su ubicación.
- COD_COM y NOM_COM: identifican la comuna de residencia del postulante.

3. C1: Postulaciones etapa regular

Esta base contiene la información correspondiente a las postulaciones realizadas por los estudiantes durante la etapa regular del Sistema de Admisión Escolar. A diferencia de la base B1, en la cual cada registro corresponde a un postulante, en esta base un mismo estudiante puede aparecer en múltiples observaciones, dado que puede incluir más de un establecimiento dentro de su listado de preferencias.

Entre las principales variables disponibles en esta base se encuentran:

- MRUN: identificador anonimizado del postulante, utilizado para vincular las postulaciones con las demás bases del proceso de admisión.
- NIVEL: identifica el nivel educacional al cual postula el estudiante.
- RBD: identifica el establecimiento educacional al cual corresponde cada postulación.
- PREFERENCIA_POSTULANTE: indica la posición que ocupa cada establecimiento dentro del orden de preferencias declarado por el postulante, donde el valor 1 corresponde a la primera preferencia.
- AGREGADA_POR_CONTINUIDAD: identifica aquellos establecimientos que fueron incorporados automáticamente por el sistema con el objetivo de asegurar la continuidad de estudios del estudiante en caso de no ser asignado a alguna de sus preferencias declaradas.

4. D1: Resultados etapa regular

La base D1 contiene los resultados del proceso de asignación correspondiente a la etapa regular del Sistema de Admisión Escolar. Esta base permite identificar, para cada postulante,

el establecimiento al que fue admitido y el nivel educacional asociado a dicha asignación.

Entre las principales variables utilizadas en esta base se encuentran:

- **MRUN:** identificador anonimizado del postulante, utilizado para vincular los resultados de asignación con las demás bases del proceso.
- **COD_NIVEL:** identifica el nivel educacional al cual postuló el estudiante.
- **RBD_ADMITIDO:** identifica el establecimiento educacional en el que el postulante fue admitido durante la etapa regular.
- **RESPUESTA_POSTULANTE:** registra la respuesta del estudiante frente a la asignación obtenida, distinguiendo entre aceptación, rechazo, permanencia en lista de espera u otras situaciones definidas por el proceso.

6.2. Consolidación de Datos

6.2.1. Fusión de las bases de datos

Para la construcción de la base utilizada en el análisis de asignación, fue necesario integrar la información proveniente de las distintas tablas del Sistema de Admisión Escolar descritas en la sección anterior. Este procedimiento se realizó de manera equivalente para cada uno de los años considerados en el estudio.

A partir de la base C1 se identificó, para cada postulante, el establecimiento declarado como primera preferencia. Para ello, se seleccionaron las observaciones en que la variable `preferencia_postulante` toma el valor 1, conservando el identificador del estudiante, el nivel de postulación y el RBD del establecimiento correspondiente a dicha preferencia.

Posteriormente, esta información fue vinculada con la base D1, la cual contiene el resultado del proceso de asignación de la etapa regular. La unión se realizó mediante el identificador `mrun`, permitiendo asociar a cada postulante el establecimiento declarado como primera preferencia y el establecimiento en el cual fue finalmente admitido.

Para el análisis geográfico, se incorporó además la información proveniente de las bases A1 y B1. La base B1 permitió identificar la ubicación aproximada del hogar de cada postulante, mientras que la base A1 proporcionó las coordenadas geográficas de los establecimientos educacionales. Estas últimas fueron vinculadas tanto al establecimiento de primera preferencia como al establecimiento asignado, utilizando el identificador `rbd` y el nivel correspondiente.

Previo a la vinculación definitiva de las bases, se verificó la unicidad de los identificadores utilizados. En particular, se comprobó que, una vez seleccionada la primera preferencia declarada en la base C1, cada combinación de estudiante y nivel educacional correspondiera a una única observación, y que en la base D1 cada identificador `mrun` estuviera asociado a un único registro. Esta revisión no evidenció observaciones duplicadas bajo dichos criterios en los años analizados.

De esta forma, se construyó una base integrada que reúne, para cada postulante, su nivel de postulación, el establecimiento declarado como primera preferencia, el establecimiento asignado y la información geográfica necesaria para los análisis posteriores.

6.2.2. Construcción de la variable de asignación

A partir de la base consolidada se construyó una variable utilizada para analizar la correspondencia entre las preferencias declaradas por las familias y los resultados obtenidos en el proceso de admisión.

En primer lugar, se identificó el establecimiento correspondiente a la primera preferen-

cia de cada postulante a partir de la base C1, seleccionando las observaciones en que la variable `preferencia_postulante` toma el valor 1. Esta información permitió obtener, para cada estudiante, el RBD del establecimiento ubicado en la primera posición de su listado de preferencias.

Posteriormente, el establecimiento de primera preferencia fue comparado con el establecimiento en el cual el estudiante fue admitido durante la etapa regular, identificado mediante la variable `rbd_admitido` de la base D1. A partir de esta comparación se construyó la variable dicotómica `admitido_primera_pref`, definida de la siguiente manera:

$$admitido_primera_pref_i = \begin{cases} 1, & \text{si } RBD_i^{primera} = RBD_i^{admitido}, \\ 0, & \text{en caso contrario.} \end{cases} \quad (5)$$

De esta forma, la variable toma el valor 1 cuando el estudiante fue admitido en el establecimiento declarado como su primera preferencia y el valor 0 cuando fue admitido en un establecimiento diferente o no presenta un establecimiento admitido en la etapa regular.

6.2.3. Construcción de la información geográfica

Para complementar el análisis de asignación con información geográfica, se construyó una base que permite identificar la ubicación aproximada del hogar de cada postulante, así como la ubicación del establecimiento declarado como primera preferencia y del establecimiento asignado en la etapa regular. Este procedimiento se realizó de manera equivalente para cada uno de los años considerados en el estudio.

En primer lugar, a partir de la base B1 se obtuvieron las coordenadas geográficas aproximadas asociadas al lugar de residencia de cada postulante, correspondientes a las variables de

latitud y longitud disponibles. Para los años 2016 y 2017, la documentación señala que, con el objetivo de resguardar la privacidad de los estudiantes, a estas coordenadas se les incorporó un error aleatorio proveniente de una distribución uniforme entre -100 y 100 metros. Para 2018 se aplicó un procedimiento similar, aunque la perturbación se encuentra expresada en grados, mediante una distribución uniforme entre $-0,0013$ y $0,0013$. Para los años 2022, 2023 y 2024 no se dispone de documentación que permita determinar si este procedimiento se mantuvo o si las coordenadas fueron sometidas a un tratamiento diferente. Por esta razón, la información geográfica utilizada se considera una aproximación a la ubicación de residencia de los postulantes y no su localización exacta.

Posteriormente, mediante la base C1 se identificó el código del establecimiento declarado como primera preferencia, mientras que a partir de la base D1 se obtuvo el código del establecimiento asignado durante la etapa regular del proceso de admisión. Esta información permitió identificar los dos establecimientos relevantes para cada postulante.

Las coordenadas geográficas de ambos establecimientos fueron obtenidas a partir de la base A1. Para ello, la información de ubicación de los establecimientos fue vinculada, mediante su respectivo código `rbd` y el nivel educacional correspondiente, tanto con el establecimiento declarado como primera preferencia como con el establecimiento asignado. De esta manera, para cada postulante fue posible disponer de tres pares de coordenadas: aquellas asociadas a su lugar de residencia aproximado, a su establecimiento de primera preferencia y al establecimiento asignado.

Finalmente, se verificó la disponibilidad de la información geográfica necesaria para los análisis posteriores. Para el análisis de distancias se consideraron únicamente aquellos postulantes para los cuales se encontraban disponibles las coordenadas de los tres puntos anteriormente señalados, permitiendo trabajar con una muestra con información geográfica completa.

6.3. Metodología

6.3.1. Análisis de asignación a primera preferencia

A partir de la variable `admitido_primera_pref`, construida previamente en la sección 6.2.2, se analizó la proporción de estudiantes que fueron admitidos en el establecimiento declarado como su primera preferencia.

El análisis se realizó para cada año y nivel educacional considerado. Para ello, se calculó el número total de postulantes y el número de estudiantes cuyo valor de `admitido_primera_pref` fue igual a 1. A partir de estos valores se obtuvo el porcentaje de estudiantes admitidos en su primera preferencia mediante la siguiente expresión:

$$P_{1ra} = \frac{N_{1ra}}{N_{post}} \times 100 \quad (6)$$

Donde N_{1ra} corresponde al número de estudiantes admitidos en su primera preferencia y N_{post} al número total de postulantes considerados en el análisis.

Los conteos se realizaron utilizando el identificador `mrún` como unidad de identificación individual de los postulantes. De esta manera, cada `mrún` distinto fue contabilizado como un estudiante dentro del nivel educacional correspondiente.

Este procedimiento permite comparar los resultados de asignación a primera preferencia entre los distintos años y niveles educacionales analizados.

6.3.2. Análisis de distancia entre residencia y establecimientos

Con el objetivo de complementar el análisis de asignación, se estimaron las distancias entre la residencia aproximada de cada postulante y los establecimientos educacionales considerados en

el proceso de admisión. En particular, se calcularon dos distancias para cada estudiante: la distancia entre su residencia y el establecimiento declarado como primera preferencia, y la distancia entre su residencia y el establecimiento en el cual fue finalmente admitido.

Dado que la información geográfica disponible en las bases del Sistema de Admisión Escolar corresponde a coordenadas de latitud y longitud, las distancias fueron calculadas mediante la fórmula de Haversine. Este método permite estimar la distancia entre dos puntos sobre la superficie terrestre a partir de sus coordenadas geográficas, considerando la curvatura de la Tierra.

La distancia de Haversine entre dos puntos se expresa como:

$$d = 2R \arcsin \left(\sqrt{\sin^2 \left(\frac{\Delta\varphi}{2} \right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2 \left(\frac{\Delta\lambda}{2} \right)} \right), \quad (7)$$

donde:

- d : distancia entre ambos puntos, expresada en kilómetros.
- R : radio medio de la Tierra, considerado igual a 6.371 km.
- φ_1 y φ_2 : latitudes de los dos puntos, expresadas en radianes.
- $\Delta\varphi$: diferencia entre las latitudes de ambos puntos.
- $\Delta\lambda$: diferencia entre las longitudes de ambos puntos.

A partir de esta expresión se construyeron las variables `distancia_primera_pref_km` y `distancia_asignado_km`, correspondientes a la distancia desde la residencia aproximada del estudiante hasta el establecimiento declarado como primera preferencia y hasta el establecimiento asignado, respectivamente.

Posteriormente, se calculó la diferencia entre ambas distancias mediante la siguiente expresión:

$$Diferencia_i = Distancia_i^{Asignado} - Distancia_i^{Primera}. \quad (8)$$

De esta manera, un valor negativo indica que el establecimiento asignado se encuentra más cerca de la residencia aproximada del postulante que su primera preferencia, mientras que un valor positivo indica que el establecimiento asignado se encuentra a una mayor distancia. Un valor igual a cero indica que ambas distancias son equivalentes. Si bien esta situación puede presentarse aun cuando la primera preferencia y el establecimiento asignado correspondan a establecimientos distintos, su ocurrencia es poco frecuente en los datos analizados.

Adicionalmente, los resultados fueron agrupados en categorías según la magnitud de la diferencia de distancia geográfica observada. En primer lugar, se distinguieron los casos en que el estudiante fue asignado al establecimiento declarado como primera preferencia y aquellos en que, pese a ser asignado a una alternativa distinta, el establecimiento obtenido se encontraba geográficamente más cerca de su residencia. Para los casos en que la asignación implicó un aumento de distancia respecto de la primera preferencia, se establecieron cuatro intervalos: menos de 0,5 km, entre 0,5 y 2 km, entre 2 y 5 km, y 5 km o más. Estos intervalos se definieron para facilitar la interpretación de resultados y distinguir entre aumentos de menor y mayor distancia. El primer intervalo, correspondiente a diferencias inferiores a 0,5 km, permite agrupar los cambios de menor magnitud, considerando además que las coordenadas de residencia utilizadas son aproximadas y que para los primeros años analizados, presentan una perturbación de 100 metros. Los intervalos siguientes permiten identificar aumentos progresivamente mayores, hasta llegar a aquellos de 5 km

o más. Estas categorías fueron definidas específicamente para este análisis y no corresponden a umbrales establecidos por el SAE.

6.3.3. Estadísticas descriptivas

A continuación, se presentan algunas estadísticas descriptivas de las bases del Sistema de Admisión Escolar utilizadas en el análisis, con el objetivo de caracterizar la muestra y entregar una visión general de la información disponible.

Cantidad de postulantes por año

La siguiente tabla presenta la cantidad de postulantes considerados en cada uno de los años analizados.

Año	N° de postulantes
2016	3.436
2017	76.821
2018	274.990
2022	570.891
2023	536.353
2024	473.482

Tabla 27: Cantidad de postulantes por año

Se puede observar un aumento importante en el tamaño de la muestra a medida que avanza la implementación del Sistema de Admisión Escolar. Este crecimiento se encuentra relacionado con la incorporación gradual de regiones y niveles educacionales al sistema. En 2016, el SAE comenzó como experiencia piloto únicamente en la Región de Magallanes y de la Antártica Chilena, considerando los niveles de ingreso de prekínder, kínder, 1° básico, 7° básico y 1° medio. En 2017, el sistema se amplió a todos los niveles en Magallanes y se incorporaron las regiones de Tarapacá, Coquimbo, O'Higgins y Los Lagos para los principales niveles de ingreso. Posteriormente, durante 2018 se incorporaron las demás regiones del país, con excepción de la Región Metropolitana, la

cual se integró en 2019, completando así la implementación territorial del sistema.

Esto se refleja en el número de postulantes disponibles en las bases analizadas, que aumenta desde 3.436 estudiantes en 2016 hasta superar los 570 mil en 2022. A partir de ese año, la muestra se mantiene considerablemente más amplia que durante las primeras etapas de implementación, aunque se observa una disminución gradual en los años posteriores, alcanzando 473.482 postulantes en 2024.

Distribución de postulantes por nivel educacional

Con el objetivo de caracterizar la composición de los postulantes considerados en el análisis, a continuación se presenta la distribución porcentual de los estudiantes según el nivel educacional al que postularon en cada año.

Nivel	2016	2017	2018	2022	2023	2024
Pre kínder	34,78 %	29,78 %	30,77 %	20,20 %	19,85 %	19,92 %
Kínder	16,65 %	12,46 %	9,57 %	6,83 %	6,71 %	6,09 %
1° Básico	13,24 %	17,13 %	14,18 %	11,43 %	10,43 %	11,12 %
2° Básico	—	0,31 %	1,54 %	4,57 %	3,74 %	2,88 %
3° Básico	—	0,22 %	1,40 %	4,40 %	4,04 %	3,24 %
4° Básico	—	0,24 %	1,26 %	3,65 %	4,06 %	3,58 %
5° Básico	—	0,21 %	1,22 %	3,71 %	3,56 %	3,91 %
6° Básico	—	0,19 %	1,08 %	3,80 %	3,68 %	3,35 %
7° Básico	5,06 %	9,08 %	7,27 %	7,01 %	7,02 %	6,84 %
8° Básico	—	0,23 %	1,14 %	4,12 %	4,39 %	4,14 %
1° Medio	30,27 %	29,85 %	28,52 %	22,64 %	23,90 %	26,33 %
2° Medio	—	0,18 %	1,07 %	4,01 %	4,72 %	4,80 %
3° Medio	—	0,11 %	0,82 %	2,96 %	3,16 %	3,12 %
4° Medio	—	0,02 %	0,14 %	0,68 %	0,75 %	0,70 %

Nota: Los porcentajes representan la proporción de postulantes de cada nivel educacional respecto del total de postulantes registrados en el SAE para cada año. El símbolo “—” indica que no se registran postulantes para dicho nivel en la base correspondiente.

Tabla 28: Distribución porcentual de postulantes al SAE según nivel educacional

La distribución por nivel educacional evidencia diferencias importantes entre los primeros años de implementación del SAE y los períodos posteriores. En 2016, la participación se concentra

principalmente en prekínder, 1° medio, kínder y 1° básico, los cuales en conjunto representan la mayor parte de los postulantes registrados durante ese año. Asimismo, 7° básico presenta una participación menor, mientras que para los restantes niveles no se observan postulaciones en la base analizada.

En 2017 y 2018 comienzan a aparecer postulantes correspondientes a una mayor cantidad de niveles educacionales, aunque la distribución continúa concentrándose principalmente en pre-kínder, 1° medio, 1° básico y kínder. Esta composición es consistente con el proceso gradual de expansión del Sistema de Admisión Escolar durante sus primeros años de funcionamiento.

En los años 2022, 2023 y 2024, en cambio, la participación se encuentra distribuida entre prácticamente todos los niveles educacionales. A pesar de ello, prekínder y 1° medio continúan concentrando una proporción relevante de las postulaciones. En 2024, por ejemplo, estos niveles representan un 19,92 % y un 26,33 % del total de postulantes, respectivamente. En contraste, 4° medio presenta una participación considerablemente menor durante los tres últimos años analizados.

Cantidad de preferencias declaradas por los postulantes

Como complemento a la caracterización de los postulantes, la siguiente tabla presenta la distribución de los estudiantes según la cantidad de establecimientos incluidos en su lista de preferencias.

Cantidad de preferencias	2016	2017	2018	2022	2023	2024
1 preferencia	0,81 %	3,70 %	4,23 %	2,98 %	2,90 %	3,21 %
2 preferencias	30,15 %	25,45 %	30,26 %	34,44 %	32,63 %	34,00 %
3 a 5 preferencias	59,25 %	59,06 %	55,21 %	52,90 %	53,96 %	50,53 %
6 o más preferencias	9,78 %	11,80 %	10,30 %	9,68 %	10,50 %	12,26 %

Tabla 29: Distribución porcentual de postulantes según cantidad de preferencias declaradas

A partir de la tabla, se puede observar un patrón estable entre los distintos años analizados. La mayor parte de los postulantes declara entre tres y cinco preferencias, categoría que concentra más de la mitad de los estudiantes en prácticamente todo el periodo.

La segunda categoría con mayor participación corresponde a los postulantes que declaran dos preferencias. Su proporción varía entre un 25,45 % en 2017 y un 34,44 % en 2022.

Por otra parte, los estudiantes que incluyen únicamente un establecimiento en su lista representan una proporción reducida en todos los años considerados, alcanzando valores inferiores al 5 %. En contraste, quienes presentan listas más extensas, correspondientes a seis o más preferencias, representan aproximadamente entre un 10 % y un 12 % de los postulantes.

6.4. Resultados

A continuación, se presentan los principales resultados obtenidos a partir del análisis del proceso de asignación del Sistema de Admisión Escolar. En primer lugar, se analiza la proporción de postulantes que fueron admitidos en el establecimiento declarado como su primera preferencia, considerando las diferencias entre niveles educacionales y años. Posteriormente, se examinan las diferencias de distancia entre la residencia aproximada de los postulantes, el establecimiento declarado como primera preferencia y el establecimiento asignado en la etapa regular.

6.4.1. Asignación al establecimiento de primera preferencia

En primer lugar, se analiza la proporción de postulantes que fueron admitidos en el establecimiento declarado como su primera preferencia. La siguiente tabla presenta los porcentajes por nivel educacional y año, permitiendo comparar la correspondencia entre las preferencias declaradas por las familias y el resultado del proceso de asignación en los distintos períodos considerados.

Nivel	2016	2017	2018	2022	2023	2024
Pre kínder	68,10 %	70,12 %	68,18 %	74,09 %	76,50 %	80,59 %
Kínder	50,70 %	49,45 %	49,15 %	50,76 %	52,66 %	58,48 %
1° Básico	45,30 %	48,51 %	52,20 %	50,49 %	55,81 %	61,41 %
2° Básico	–	29,58 %	22,56 %	23,05 %	27,80 %	33,84 %
3° Básico	–	24,26 %	23,31 %	22,78 %	24,84 %	30,55 %
4° Básico	–	22,53 %	25,17 %	26,26 %	20,95 %	23,89 %
5° Básico	–	41,36 %	25,43 %	25,31 %	26,34 %	23,35 %
6° Básico	–	35,66 %	25,69 %	21,47 %	22,45 %	23,92 %
7° Básico	47,10 %	59,63 %	59,67 %	43,19 %	43,24 %	46,89 %
8° Básico	–	28,00 %	32,01 %	22,92 %	19,42 %	20,80 %
1° Medio	61,30 %	60,30 %	65,06 %	52,18 %	51,20 %	54,04 %
2° Medio	–	35,04 %	54,91 %	33,11 %	30,42 %	31,04 %
3° Medio	–	44,44 %	47,35 %	39,13 %	41,79 %	40,62 %
4° Medio	–	50,00 %	80,81 %	63,35 %	61,55 %	59,85 %

Tabla 30: Porcentaje de postulantes admitidos en su primera preferencia según nivel educacional

Los resultados muestran diferencias relevantes en la proporción de estudiantes admitidos en su primera preferencia según el nivel educacional y el año analizado. En términos generales, los niveles de ingreso al sistema escolar presentan porcentajes más elevados de asignación a primera preferencia. En particular, prekínder registra los valores más altos durante todo el período considerado, pasando de un 68,10 % en 2016 a un 80,59 % en 2024. Asimismo, kínder y 1° básico presentan porcentajes relativamente altos, especialmente en los años más recientes, alcanzando en 2024 un 58,48 % y un 61,41 %, respectivamente.

Un comportamiento similar se observa en 1° medio, nivel que presenta porcentajes superiores al 50 % en todos los años analizados. Durante los primeros periodos, la proporción de estudiantes admitidos en su primera preferencia alcanza un 61,30 % en 2016 y un 65,06 % en 2018, mientras que en los años posteriores se sitúa en torno al 51 % y 54 %.

En contraste, varios de los niveles intermedios de educación básica presentan porcentajes considerablemente menores. Entre 2° y 8° básico, excluyendo 7° básico, las tasas de asignación a

primera preferencia se sitúan generalmente entre aproximadamente un 20 % y un 35 % durante los años en que existen registros. Dentro de este grupo, 4°, 5°, 6° y 8° básico presentan algunos de los porcentajes más bajos del período. Por ejemplo, en 2024 la proporción alcanza un 23,89 % en 4° básico, un 23,35 % en 5° básico, un 23,92 % en 6° básico y un 20,80 % en 8° básico.

Por su parte, 7° básico presenta un comportamiento distinto al resto de los niveles intermedios, con porcentajes más elevados, especialmente en 2017 y 2018, cuando aproximadamente un 60 % de los postulantes fue admitido en su primera preferencia. En los años 2022, 2023 y 2024 esta proporción disminuye, situándose entre un 43 % y un 47 %.

En enseñanza media también se observan diferencias entre niveles. Mientras 1° medio presenta una proporción relativamente alta de asignación a primera preferencia, 2° y 3° medio registran porcentajes menores. En cambio, 4° medio exhibe valores elevados, aunque este resultado debe interpretarse considerando que dicho nivel representa una proporción muy reducida del total de postulantes, como se observó previamente en las estadísticas descriptivas.

En conjunto, los resultados evidencian que la probabilidad de obtener el establecimiento declarado como primera preferencia no es homogénea entre los distintos niveles educacionales. Los mayores porcentajes tienden a concentrarse en algunos niveles de ingreso, como prekínder, 1° básico y 1° medio, mientras que los niveles intermedios presentan, en general, menores tasas de asignación a primera preferencia.

6.4.2. Distancia entre el establecimiento asignado y la primera preferencia

Con el objetivo de complementar el análisis de asignación, se examina la diferencia entre la distancia desde la residencia aproximada del postulante hasta el establecimiento declarado como primera preferencia y la distancia hasta el establecimiento asignado en la etapa regular. Para faci-

litar la interpretación, los resultados fueron agrupados en categorías según la magnitud y dirección de esta diferencia. La siguiente tabla presenta la distribución porcentual de los postulantes en cada una de estas categorías para los años analizados.

Resultado de distancia	2016	2017	2018	2022	2023	2024
Asignado a 1ra preferencia	59,08 %	64,93 %	64,80 %	52,29 %	52,81 %	56,55 %
Asignado más cerca	23,11 %	15,99 %	16,35 %	20,07 %	19,67 %	18,06 %
Cambio menor a 0,5 km	7,77 %	6,23 %	5,74 %	7,66 %	7,69 %	7,08 %
Aumento entre 0,5–2 km	7,65 %	7,14 %	7,43 %	10,35 %	10,29 %	9,45 %
Aumento entre 2–5 km	2,21 %	3,29 %	3,27 %	4,93 %	4,82 %	4,52 %
Aumento igual o mayor a 5 km	0,17 %	2,42 %	2,41 %	4,70 %	4,73 %	4,34 %

Tabla 31: Distribución porcentual de postulantes según diferencia de distancia entre la primera preferencia y el establecimiento asignado

Los resultados muestran que, en todos los años considerados, la categoría con mayor porcentaje corresponde a los estudiantes que fueron asignados al establecimiento declarado como su primera preferencia. Esta proporción alcanza un 59,08 % en 2016 y aumenta a aproximadamente un 65 % en 2017 y 2018. En los años posteriores, disminuye hasta valores cercanos al 52 % en 2022 y 2023, para luego aumentar a un 56,55 % en 2024.

Entre los estudiantes que no fueron asignados a su primera preferencia, una proporción relevante fue admitida en un establecimiento ubicado más cerca de su residencia. Esta categoría representa un 23,11 % en 2016 y posteriormente se mantiene entre aproximadamente un 16 % y un 20 % durante los restantes años. Por lo tanto, no todos los casos en que el estudiante es asignado a una alternativa distinta de su primera preferencia implican necesariamente un aumento en la distancia respecto de su residencia.

En cuanto a los postulantes cuya asignación implicó un aumento de distancia, los cambios inferiores a 0,5 km representan entre un 5,74 % y un 7,77 % de los casos. Los aumentos correspondientes a diferencias entre 0,5 y 2 km, presentan una participación algo mayor en los años

recientes, pasando de alrededor de un 7 % durante 2016–2018 a aproximadamente un 10 % entre 2022 y 2024.

Por su parte, los aumentos superiores a 2 km representan una proporción menor de los postulantes, aunque su presencia es mayor en los últimos años analizados. En particular, la categoría de aumento entre 2 y 5 km pasa de un 2,21 % en 2016 a valores cercanos al 5 % en 2022 y 2023. Del mismo modo, los aumentos iguales o superiores a 5 km representan apenas un 0,17 % en 2016, pero alcanzan aproximadamente un 4,7 % en 2022 y 2023 y un 4,34 % en 2024.

En conjunto, los resultados indican que la mayoría de los postulantes analizados fue asignada a su primera preferencia o a un establecimiento ubicado más cerca de su residencia. Sin embargo, también se observa que en los años más recientes aumenta la proporción de estudiantes cuya asignación implica desplazamientos adicionales de mayor magnitud respecto del establecimiento declarado como primera preferencia.

7. Conclusiones

La presente investigación tuvo como objetivo analizar distintas dimensiones asociadas al funcionamiento del Sistema de Admisión Escolar en Chile, considerando su relación con la segregación socioeconómica, la evolución del rendimiento académico y la correspondencia entre las preferencias declaradas por las familias y los resultados de asignación obtenidos. Para ello, se utilizaron bases SIMCE y registros administrativos del SAE, construyendo distintas cohortes de estudiantes y aplicando herramientas descriptivas y econométricas que permitieron estudiar estos fenómenos desde una perspectiva longitudinal y territorial.

En primer lugar, respecto de la segregación socioeconómica, los resultados obtenidos mediante el índice de Duncan muestran que existe una reducción generalizada de la segregación entre 4° básico y 2° medio en prácticamente todas las regiones y cohortes analizadas. Sin embargo, este comportamiento no se observa únicamente en las cohortes posteriores a la implementación del SAE, sino también en aquellas previas al sistema. Las cohortes pre-SAE presentan variaciones promedio negativas del índice de Duncan de magnitud similar a las observadas en las cohortes post-SAE, lo que indica que la disminución de la segregación entre ambos niveles educativos constituye un patrón que ya se encontraba presente antes de la implementación del nuevo mecanismo de admisión.

Los análisis estadísticos refuerzan esta conclusión. Tanto en las cohortes previas como posteriores al SAE, las variaciones promedio del índice de Duncan resultan negativas y estadísticamente significativas. Además, si bien durante el período post-SAE una mayor cantidad de regiones presenta disminuciones estadísticamente significativas, las reducciones observadas no son sistemáticamente mayores que aquellas registradas en el período previo. Por lo tanto, a partir de los

resultados obtenidos no es posible sostener que la reducción observada en la segregación socioeconómica pueda atribuirse exclusivamente a la implementación del Sistema de Admisión Escolar. Más bien, los resultados sugieren que existen dinámicas asociadas a la trayectoria educativa y a la estructura territorial del sistema escolar que también inciden en la evolución de la segregación.

Asimismo, el análisis territorial muestra que los niveles de segregación socioeconómica presentan una importante heterogeneidad entre regiones. Algunas zonas del país, especialmente aquellas con una mayor proporción de estudiantes de bajo nivel socioeconómico y una mayor presencia de establecimientos rurales, presentan niveles de segregación relativamente superiores. Del mismo modo, se observan diferencias relevantes en la disponibilidad de establecimientos educacionales entre regiones y niveles de enseñanza. En particular, la menor oferta de establecimientos en educación media respecto de educación básica constituye un elemento que puede restringir las posibilidades de elección de las familias y afectar la distribución de los estudiantes entre establecimientos.

En segundo lugar, el análisis del rendimiento académico mediante modelos de Diferencias en Diferencias muestra que las cohortes posteriores a la implementación del SAE presentan una evolución menos favorable de los puntajes SIMCE en comparación con las cohortes previas. En Lectura, el coeficiente asociado al estimador DiD es negativo y estadísticamente significativo en las tres especificaciones, con una magnitud cercana a 1,5 puntos en el modelo más completo. En Matemática, la diferencia estimada es considerablemente mayor, alcanzando aproximadamente 14,9 puntos negativos una vez incorporados los controles socioeconómicos, la dependencia administrativa, el sexo del estudiante y los efectos fijos por comuna.

La estabilidad del coeficiente DiD frente a la incorporación progresiva de controles muestra que las diferencias observadas entre las cohortes pre-SAE y post-SAE no se explican completa-

mente por las características socioeconómicas, administrativas o territoriales incorporadas en los modelos. No obstante, estos resultados deben ser interpretados como diferencias en la evolución de los puntajes entre ambos grupos de cohortes y no como evidencia suficiente para atribuir de manera exclusiva dichos cambios al funcionamiento del Sistema de Admisión Escolar.

Los modelos también evidencian la persistencia de importantes brechas asociadas a las características de los estudiantes y de los establecimientos. En ambas pruebas, los estudiantes clasificados como de nivel socioeconómico bajo presentan puntajes significativamente menores que aquellos pertenecientes al grupo socioeconómico alto. Estas diferencias son especialmente pronunciadas en Matemática. Asimismo, se observan diferencias según dependencia administrativa y sexo del estudiante, confirmando que el rendimiento académico continúa estrechamente relacionado con características individuales y del entorno educativo.

Finalmente, el análisis de las preferencias y resultados de asignación permite observar que la capacidad del Sistema de Admisión Escolar para asignar a los estudiantes al establecimiento declarado como su primera preferencia varía de manera importante entre niveles educacionales y años. Los mayores porcentajes de asignación a primera preferencia se concentran principalmente en niveles de ingreso al sistema, como prekínder, 1° básico y 1° medio, mientras que varios niveles intermedios de educación básica presentan tasas considerablemente menores.

El análisis geográfico complementa estos resultados al mostrar que no todos los estudiantes que no obtienen su primera preferencia son necesariamente asignados a un establecimiento más alejado de su residencia. En todos los años analizados, una proporción relevante de postulantes fue asignada a un establecimiento ubicado más cerca de su hogar que aquel declarado como primera preferencia. Sin embargo, también existe un grupo de estudiantes cuya asignación implica un aumento de distancia, y esta situación adquiere mayor relevancia en los años más recientes.

En conjunto, los resultados de esta investigación muestran que la implementación del Sistema de Admisión Escolar introdujo un cambio importante en las reglas de acceso a los establecimientos educacionales, reemplazando mecanismos de selección descentralizados por un proceso centralizado y estandarizado. Sin embargo, los resultados obtenidos también muestran que la centralización del proceso de admisión no elimina por sí sola las desigualdades existentes en el sistema escolar.

La reducción de la segregación socioeconómica observada en las cohortes post-SAE también se encontraba presente antes de su implementación, mientras que las diferencias en rendimiento académico continúan fuertemente asociadas a características socioeconómicas y al tipo de establecimiento. Asimismo, la correspondencia entre las preferencias de las familias y los resultados de asignación varía considerablemente según el nivel educacional, y para una parte de los postulantes la asignación final puede implicar una distancia mayor respecto de su residencia.

Por lo tanto, los resultados sugieren que el Sistema de Admisión Escolar debe entenderse como un mecanismo que modifica las reglas mediante las cuales se distribuyen los estudiantes entre establecimientos, pero cuyo funcionamiento se encuentra condicionado por factores estructurales que van más allá del procedimiento de asignación. La composición socioeconómica de los territorios, la distribución de la oferta educativa, las preferencias de las familias y las diferencias existentes entre establecimientos continúan desempeñando un papel relevante en las oportunidades educativas de los estudiantes. En este sentido, avanzar hacia una mayor equidad educativa requiere considerar no solo el diseño del mecanismo de admisión, sino también las condiciones estructurales del sistema escolar en las cuales dicho mecanismo opera.

8. Limitaciones

La investigación presenta algunas limitaciones que deben ser consideradas al momento de interpretar los resultados obtenidos. Estas se relacionan principalmente con la disponibilidad y características de las bases de datos utilizadas, la construcción de las cohortes analizadas y las condiciones particulares de los períodos considerados.

En primer lugar, el análisis de segregación socioeconómica se encuentra condicionado por la disponibilidad de información sobre los ingresos de los hogares en las bases SIMCE. La clasificación socioeconómica utilizada se construyó a partir del ingreso reportado por padres y apoderados en los cuestionarios asociados a esta evaluación, por lo que corresponde a información autodeclarada y agrupada en tramos de ingreso. Además, la estructura de esta variable presenta diferencias entre años, lo que hizo necesario realizar procesos de homologación para mantener un criterio comparable durante todo el período analizado. Asimismo, aquellos estudiantes que no contaban con información válida de ingresos fueron excluidos del análisis, por lo que la muestra utilizada para calcular el índice de Duncan no corresponde a la totalidad de los estudiantes matriculados en cada año.

Adicionalmente, la clasificación socioeconómica empleada divide a los estudiantes en dos grupos, alto y bajo nivel socioeconómico, utilizando como referencia el tramo de ingreso exento del Impuesto a la Renta. Si bien este criterio permite construir una clasificación consistente entre años y aplicar el índice de Duncan, simplifica la heterogeneidad existente entre los hogares y no incorpora otras dimensiones socioeconómicas, tales como el nivel educacional de los padres, condiciones de vivienda u otras características que podrían contribuir a describir con mayor precisión la situación socioeconómica.

Una segunda consideración se relaciona con la comparación del índice de Duncan entre 4° básico y 2° medio. Los resultados muestran una reducción de la segregación socioeconómica entre ambos niveles tanto en las cohortes previas como posteriores a la implementación del SAE. Sin embargo, parte de esta variación puede estar asociada a cambios propios de la trayectoria educativa de los estudiantes y de la oferta de establecimientos disponible en cada nivel. Como se observó en las estadísticas descriptivas, la cantidad de establecimientos y la presencia de establecimientos rurales disminuyen considerablemente al pasar desde educación básica hacia educación media. Por lo tanto, las variaciones del índice no necesariamente responden únicamente a cambios en los mecanismos de admisión.

En relación con el análisis del rendimiento académico, una de las principales limitaciones corresponde a la presencia de acontecimientos externos que coinciden temporalmente con las cohortes consideradas. En particular, las cohortes posteriores a la implementación del SAE estuvieron expuestas a la pandemia de COVID-19, al cierre prolongado de los establecimientos educacionales y al posterior proceso de retorno a la presencialidad. Estos acontecimientos pudieron afectar de manera significativa la trayectoria de los aprendizajes, por lo que los coeficientes obtenidos mediante el modelo de Diferencias en Diferencias no pueden interpretarse como consecuencias atribuibles exclusivamente a la implementación del Sistema de Admisión Escolar.

De manera similar, las cohortes utilizadas como grupo de comparación tampoco se encuentran completamente libres de acontecimientos particulares. En especial, parte de los estudiantes pertenecientes a las cohortes previas fue evaluada en 4° básico durante 2010, año marcado por el terremoto ocurrido en Chile. En las zonas más afectadas, este evento pudo alterar temporalmente el funcionamiento de los establecimientos y el desempeño de los estudiantes. En consecuencia, las diferencias observadas entre cohortes pueden incorporar efectos asociados a acontecimientos

distintos del SAE, dificultando la identificación aislada del impacto del sistema de admisión.

Asimismo, la estrategia de Diferencias en Diferencias depende del supuesto de que, en ausencia del tratamiento, las cohortes pre-SAE y post-SAE habrían presentado trayectorias comparables. La disponibilidad de información utilizada en esta investigación permite observar dos puntos de la trayectoria de cada estudiante, correspondientes a 4° básico y 2° medio, pero no una serie extensa de períodos previos que permita evaluar de manera más completa la existencia de tendencias paralelas antes de la implementación del sistema. Por esta razón, aunque el modelo permite estimar diferencias en la evolución de los puntajes entre grupos y los resultados se mantienen estables ante la incorporación de distintos controles, la interpretación causal del coeficiente DiD debe realizarse con cautela.

Otra limitación del análisis de rendimiento se relaciona con la construcción de las cohortes longitudinales. Para vincular a los estudiantes entre 4° básico y 2° medio fue necesario disponer de un identificador `mrun` válido y único en ambos períodos, además de contar con puntajes SIMCE válidos en las respectivas pruebas. Como consecuencia, quedan excluidos aquellos estudiantes que no pudieron ser identificados longitudinalmente, que no participaron de alguna de las evaluaciones o que presentaban información incompleta. Por lo tanto, los resultados obtenidos corresponden sólo a los estudiantes que pudieron ser seguidos entre ambos niveles educativos.

Respecto del análisis de asignación escolar, debe considerarse que durante los primeros años el Sistema de Admisión Escolar se encontraba en un proceso de implementación gradual. En 2016 el sistema operaba únicamente en la Región de Magallanes y de la Antártica Chilena y en determinados niveles educacionales, mientras que posteriormente fueron incorporándose nuevas regiones y niveles hasta alcanzar una cobertura territorial más amplia. En consecuencia, los resultados correspondientes a 2016, 2017 y 2018 no son directamente comparables con los años

posteriores en términos de cobertura nacional, ya que representan etapas distintas del proceso de expansión del SAE. Las diferencias observadas entre años pueden reflejar, al menos parcialmente, cambios en la composición de los postulantes y de los territorios incorporados al sistema.

Además, el análisis de primera preferencia considera los resultados correspondientes a la etapa regular del proceso de admisión. De esta forma, no se incorporan los cambios que pueden producirse posteriormente en la etapa complementaria ni otras decisiones adoptadas por las familias después de conocer el resultado inicial de asignación. Por lo tanto, el indicador utilizado representa la correspondencia entre la primera preferencia declarada y la asignación registrada durante la etapa regular, y no necesariamente la situación definitiva del estudiante al momento de matricularse.

En cuanto al análisis geográfico, las coordenadas correspondientes a la residencia de los postulantes contienen un error aleatorio incorporado deliberadamente en las bases del SAE con el objetivo de resguardar la privacidad de los estudiantes. Por esta razón, las distancias estimadas representan aproximaciones y no corresponden a la distancia exacta entre la vivienda del postulante y los establecimientos educacionales. Esta imprecisión puede ser particularmente relevante cuando se analizan diferencias pequeñas entre establecimientos cercanos.

Asimismo, las distancias fueron calculadas mediante la fórmula de Haversine, la cual estima la distancia entre dos puntos a partir de sus coordenadas geográficas. Si bien este procedimiento permite aplicar un criterio uniforme para todos los postulantes, no representa necesariamente la distancia efectivamente recorrida por las familias. Factores como la red vial disponible, las rutas de transporte público, tiempos de viaje y costos de desplazamiento no son considerados en esta medida. En consecuencia, una diferencia determinada en kilómetros puede representar condiciones de accesibilidad distintas dependiendo del territorio en que se encuentre el estudiante.

Por último, el análisis de asignación y distancia se centra en identificar la correspondencia entre la primera preferencia y el establecimiento asignado, pero no permite determinar completamente las razones que explican las preferencias declaradas por las familias. La elección de un establecimiento puede responder simultáneamente a factores como su ubicación, desempeño académico, proyecto educativo, composición social, disponibilidad de cupos o características particulares del hogar. Dado que estas dimensiones no se observan de manera conjunta en las bases utilizadas para este análisis, los resultados permiten describir las diferencias entre preferencias y asignaciones, pero no identificar de manera causal los factores que explican dichas decisiones.

En conjunto, estas limitaciones no invalidan los resultados obtenidos, pero sí deben ser consideradas al interpretar su alcance. Los análisis desarrollados permiten identificar patrones relevantes en la evolución de la segregación socioeconómica, el rendimiento académico y los resultados de asignación escolar; sin embargo, la complejidad del sistema educativo y la coexistencia de distintos factores institucionales, territoriales y temporales dificultan atribuir los cambios observados exclusivamente al Sistema de Admisión Escolar.

9. Anexos

9.1. Anexo A

DiD: Resultados primera regresión

Tabla 32: Resultados de la regresión para Lectura - Variables Principales

Variable	Estimador	Error Estándar	Valor t	P-Valor
Intercepto	279.36870	0.099121	2818.4704	$< 2.2 \times 10^{-16}***$
Tratamiento	-3.85597	0.131435	-29.3374	$< 2.2 \times 10^{-16}***$
Post	-20.57010	0.100081	-205.5347	$< 2.2 \times 10^{-16}***$
Interacción DID	-1.69268	0.125613	-13.4754	$< 2.2 \times 10^{-16}***$

Observaciones: 1.151.674. RMSE: 50,8. R^2 ajustado: 0,045197.

Errores estándar agrupados a nivel de estudiante (MRUN).

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Tabla 33: Resultados de la regresión para Matemática - Variables Principales

Variable	Estimador	Error Estándar	Valor t	P-Valor
Intercepto	267.282947	0.104086	2567.91513	$< 2.2 \times 10^{-16}***$
Tratamiento	0.193601	0.129973	1.48954	0.13635
Post	10.777199	0.102685	104.95447	$< 2.2 \times 10^{-16}***$
Interacción DID	-15.194645	0.133899	-113.47861	$< 2.2 \times 10^{-16}***$

Observaciones: 1.167.118. RMSE: 56,1. R^2 ajustado: 0,008552.

Errores estándar agrupados a nivel de estudiante (MRUN).

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

9.2. Anexo B

DiD: Resultados segunda regresión

Tabla 34: Resultados de la regresión para Lectura - Variables Principales + Controles

Variable	Estimador	Error Estándar	Valor t	P-Valor
Intercepto	275.49315	0.180575	1525.64029	$< 2.2 \times 10^{-16}***$
Tratamiento	-4.71813	0.127138	-37.11015	$< 2.2 \times 10^{-16}***$
Post	-20.57010	0.100081	-205.53424	$< 2.2 \times 10^{-16}***$
Interacción DID	-1.69268	0.125613	-13.47537	$< 2.2 \times 10^{-16}***$
SES bajo	-15.77018	0.135658	-116.24930	$< 2.2 \times 10^{-16}***$
Particular Subvencionado	11.72396	0.127988	91.60197	$< 2.2 \times 10^{-16}***$
Particular Pagado	30.63754	0.233507	131.20626	$< 2.2 \times 10^{-16}***$
Servicios Locales de Educación Pública	2.57532	1.031392	2.49694	0.012527*
Mujer	10.76212	0.113600	94.73713	$< 2.2 \times 10^{-16}***$

Observaciones: 1.151.674. RMSE: 48,7. R^2 ajustado: 0,123569.

Errores estándar agrupados a nivel de estudiante (MRUN).

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Tabla 35: Resultados de la regresión para Matemática - Variables Principales + Controles

Variable	Estimador	Error Estándar	Valor t	P-Valor
Intercepto	271.30402	0.191031	1420.20989	$< 2.2 \times 10^{-16}***$
Tratamiento	-1.22578	0.124189	-9.87030	$< 2.2 \times 10^{-16}***$
Post	10.77720	0.102685	104.95425	$< 2.2 \times 10^{-16}***$
Interacción DID	-15.19464	0.133899	-113.47837	$< 2.2 \times 10^{-16}***$
SES bajo	-20.24753	0.146227	-138.46665	$< 2.2 \times 10^{-16}***$
Particular Subvencionado	17.26283	0.134831	128.03276	$< 2.2 \times 10^{-16}***$
Particular Pagado	49.57858	0.253776	195.36360	$< 2.2 \times 10^{-16}***$
Servicios Locales de Educación Pública	4.36801	1.015996	4.29924	$1.7141 \times 10^{-5}***$
Mujer	-7.02214	0.120535	-58.25793	$< 2.2 \times 10^{-16}***$

Observaciones: 1.167.118. RMSE: 52,4. R^2 ajustado: 0,135129.

Errores estándar agrupados a nivel de estudiante (MRUN).

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

9.3. Anexo C

DiD: Resultados tercera regresión

Tabla 36: Resultados de la regresión para Lectura - Variables Principales + Controles + Efectos Fijos por Comuna

Variable	Estimador	Error Estándar	Valor t	P-Valor
Tratamiento	-4.67635	0.126601	-36.93761	$< 2.2 \times 10^{-16}***$
Post	-21.37681	0.101623	-210.35489	$< 2.2 \times 10^{-16}***$
Interacción DID	-1.55218	0.126127	-12.30648	$< 2.2 \times 10^{-16}***$
SES bajo	-15.32374	0.138202	-110.87914	$< 2.2 \times 10^{-16}***$
Particular Subvencionado	12.57960	0.136291	92.29965	$< 2.2 \times 10^{-16}***$
Particular Pagado	29.11887	0.256293	113.61562	$< 2.2 \times 10^{-16}***$
Servicios Locales de Educación Pública	4.63479	1.053169	4.40081	$1.0787 \times 10^{-5}***$
Mujer	10.68923	0.112738	94.81474	$< 2.2 \times 10^{-16}***$

Observaciones: 1.151.674. Efectos fijos por comuna: 343.
 RMSE: 48,4. R^2 ajustado: 0,13408. R^2 within: 0,112849.
 Errores estándar agrupados a nivel de estudiante (MRUN).
 *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Tabla 37: Resultados de la regresión para Matemática - Variables Principales + Controles + Efectos Fijos por Comuna

Variable	Estimador	Error Estándar	Valor t	P-Valor
Tratamiento	-1.13412	0.123362	-9.19346	$< 2.2 \times 10^{-16}***$
Post	9.67919	0.104499	92.62491	$< 2.2 \times 10^{-16}***$
Interacción DID	-14.91423	0.134495	-110.89067	$< 2.2 \times 10^{-16}***$
SES bajo	-19.30655	0.147477	-130.91076	$< 2.2 \times 10^{-16}***$
Particular Subvencionado	17.31772	0.142618	121.42763	$< 2.2 \times 10^{-16}***$
Particular Pagado	44.71860	0.278891	160.34458	$< 2.2 \times 10^{-16}***$
Servicios Locales de Educación Pública	7.29429	1.032602	7.06399	$1.6197 \times 10^{-12}***$
Mujer	-7.11056	0.118649	-59.92932	$< 2.2 \times 10^{-16}***$

Observaciones: 1.167.118. Efectos fijos por comuna: 343.
 RMSE: 51,7. R^2 ajustado: 0,155557. R^2 within: 0,101313.
 Errores estándar agrupados a nivel de estudiante (MRUN).
 *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Referencias

- Abdulkadiroğlu, A., & Sönmez, T. (2003). School Choice: A Mechanism Design Approach. *American Economic Review*, 93(3), 729-747. <https://doi.org/10.1257/000282803322157061>
- Behrman, J. R., Tincani, M. M., Todd, P. E., & Wolpin, K. I. (2016). Teacher Quality in Public and Private Schools under a Voucher System: The Case of Chile. *Journal of Labor Economics*, 34(2), 319-362. <https://doi.org/10.1086/683642>
- Bellei, C., Contreras, M., & Guajardo, F. (2025). Families' responses to a new centralized school admission system. School choice and justice in education. *International Journal of Educational Development*. <https://doi.org/10.1016/j.ijedudev.2024.103202>
- Bravo, D., Mukhopadhyay, S., & Todd, P. E. (2010). Effects of school reform on education and labor market performance: Evidence from Chile's universal voucher system. *Quantitative Economics*, 1(1), 47-95. <https://doi.org/https://doi.org/10.3982/QE16>
- Burgess, S., Greaves, E., Vignoles, A., & Wilson, D. (2014). What Parents Want: School Preferences and School Choice. *The Economic Journal*, 125(587), 1262-1289. <https://doi.org/10.1111/ecoj.12153>
- Chumacero, R. A., Gómez, D., & Paredes, R. D. (2011). I would walk 500 miles (if it paid): Vouchers and school choice in Chile. *Economics of Education Review*, 30(5), 1103-1114. <https://doi.org/https://doi.org/10.1016/j.econedurev.2011.05.015>
- Decreto N°257. (2009). Biblioteca del Congreso Nacional de Chile [Santiago de Chile, 19 de agosto de 2009].

- Duncan, O. D., & Duncan, B. (1955). A Methodological Analysis of Segregation Indexes. *American Sociological Review*, 20(2), 210-217. Consultado el 19 de noviembre de 2025, desde <http://www.jstor.org/stable/2088328>
- Fredriksson, A., & de Oliveira, G. M. (2019). Impact evaluation using Difference-in-Differences. *RAUSP Management Journal*. <https://doi.org/10.1108/rausp-05-2019-0112>
- Gale, D., & Shapley, L. S. (1962). College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 69(1), 9-15. <https://doi.org/10.1080/00029890.1962.11989827>
- Gazmuri, A. M. (2024). School segregation in the presence of student sorting and cream-skimming: Evidence from a school voucher reform. *Journal of Public Economics*, 238, 105176. <https://doi.org/https://doi.org/10.1016/j.jpubeco.2024.105176>
- Godoy, F., Salazar, F., & Treviño, E. (2014). *Prácticas de selección en el sistema escolar chileno: tipos de requisitos de postulación, legitimación en el sistema y las debilidades de la ley* (inf. téc.). Centro de Políticas Comparadas de Educación, Universidad Diego Portales. Santiago, Chile. Consultado el 24 de noviembre de 2025, desde <https://www.cpce.cl/ipe>
- Greaves, E. (2024). Segregation by choice? School choice and segregation in England. *Education Economics*, 32(1), 1-19. <https://doi.org/10.1080/09645292.2023.2181748>
- Gutiérrez, G., & Carrasco, A. (2021). Chile's enduring educational segregation: A trend unchanged by different cycles of reform. *British Educational Research Journal*, 47(6), 1611-1634. <https://doi.org/https://doi.org/10.1002/berj.3746>
- Hsieh, C.-T., & Urquiola, M. (2006). The effects of generalized school choice on achievement and stratification: Evidence from Chile's voucher program. *Journal of Public Economics*, 90(8), 1477-1503. <https://doi.org/https://doi.org/10.1016/j.jpubeco.2005.11.002>

- Kutscher, M., Nath, S., & Urzúa, S. (2023). Centralized admission systems and school segregation: Evidence from a national reform. *Journal of Public Economics*, 221, 104863. <https://doi.org/https://doi.org/10.1016/j.jpubeco.2023.104863>
- Ley N°21.040. (2017). Diario Oficial de la República de Chile. [Santiago de Chile, 24 de noviembre de 2017].
- Libertad y Desarrollo. (2024). *Financiamiento compartido: Evolución, efectos y discusión actual* (Temas Públicos N.º 1656). Libertad y Desarrollo. https://lyd.org/wp-content/uploads/2024/09/TP_1656_FINANCIAMIENTO_COMPARTIDO.pdf
- MINEDUC. (s.f.-a). Planes y programas de estudio [Recuperado el 31 de octubre de 2025 de <https://www.ayudamineduc.cl/ficha/tipos-de-establecimiento-segun-financiamiento-4>].
- MINEDUC. (s.f.-b). Subsecretaría de Educación Parvularia [Recuperado el 03 de noviembre de 2025 de <https://www.ayudamineduc.cl/ficha/subsecretaria-de-educacion-parvularia>].
- Ministerio de Educación, Centro de Estudios. (2025). Análisis de la matrícula escolar en Chile año 2025. *Apuntes 97*.
- Mizala, A., & Torche, F. (2012). Bringing the schools back in: the stratification of educational achievement in the Chilean voucher system. *International Journal of Educational Development*, 32(1), 132-144. <https://doi.org/https://doi.org/10.1016/j.ijedudev.2010.09.004>
- Morgan, B. (1980). Occupational Segregation in Metropolitan Areas in the United States, 1970. *Urban Studies*, 17, 63-69. <https://doi.org/10.1080/00420988020080071>
- Oyarzún, J., Parcerisa, L., & Rozas, A. C. (2023). New and old educational inequalities in socio-cultural minorities: exploring the school choice experiences of families under the new school admission system in Chile. *Research Papers in Education*, 39, 608-628. <https://doi.org/10.1080/02671522.2023.2188246>

- Rothbard, S., Etheridge, J. C., & Murray, E. J. (2023). A Tutorial on Applying the Difference-in-Differences Method to Health Data. *Current Epidemiology Reports*. <https://doi.org/10.1007/s40471-023-00327-x>
- Santos, H., & Elacqua, G. (2017). Socioeconomic school segregation in Chile: Parental choice and a theoretical counterfactual analysis. *CEPAL Review*, 2016, 123-137. <https://doi.org/10.18356/938a3490-en>
- Unidad de Currículum y Evaluación. (2021, Abril). *Plan de Evaluaciones Nacionales e Internacionales 2021-2026* (inf. téc.). Ministerio de Educación. Santiago, Chile. https://www.curriculumnacional.cl/614/articles-236337_archivo_01.pdf