



**UNIVERSIDAD DE CHILE**

**FACULTAD DE CIENCIAS SOCIALES**

**ESCUELA DE PREGRADO - CARRERA DE SOCIOLOGÍA**

**Desigualdades de género en trayectorias  
estudiantiles en la educación terciaria chilena 2008-2021**

Memoria de Título para optar al Título Profesional de Sociólogo

Autor(a):

**Alexis Villanueva Villanueva**

Profesor(a) guía:

**Patricio Andrés Rodríguez Valdés**

Santiago de Chile

Diciembre 2022



## **Agradecimientos**

Gracias al Centro de Estudios del Ministerio de Educación, a la Agencia de Calidad en la Educación, a la Comisión Ingresos y al Departamento de Evaluación, Medición y Registro Educativo de la Universidad de Chile. Entidades que proveyeron los datos empleados en el presente trabajo.

Gracias al respaldo institucional del Centro de Investigación Avanzada en Educación de la Universidad de Chile en la solicitud de estos datos.

Se agradece al financiamiento provisto por ANID/PIA/Fondos basales para centros de excelencia FB003.

## Resumen

Se propuso investigar, empleando datos de fuentes secundarias, cuasi-censales y de carácter administrativo, las desigualdades de género en las trayectorias estudiantiles en la educación terciaria en Chile para las cohortes de egreso escolar entre 2008 y 2014.

Los antecedentes teóricos y empíricos sugieren la existencia de brechas vocacionales en la postulación e ingreso a las carreras y de aspectos que reducen la retención de las estudiantes al interior de los programas, sumado a eventuales intersecciones con otras brechas relativas a la institución o al estudiante.

Mediante una aproximación de análisis exploratorio de datos guiado por modelos se evaluó la relevancia del género en cinco hitos relativos a la trayectoria: El egreso escolar, la postulación, la primera matrícula, la retención y el eventual reingreso.

Usando valores SHAP y el índice de Shapley-Taylor se realizó una revisión exploratoria cuyos resultados son que el género es un elemento relevante en el ingreso y la postulación hacia áreas GEEMP (en inglés, geociencias, ingenierías, economía, matemáticas e informática y ciencias físicas), así como también en la retención y, eventualmente, en el reingreso hacia esta área, sin embargo, no se encontró evidencia que indique que dicha característica sea relevante en otros hitos ni que tenga efectos de interacción relevantes que den cuenta de intersección con otras características.

***Palabras clave:*** *Trayectorias estudiantiles, brechas de género, análisis exploratorio de datos.*

# Índice

1	Introducción.....	7
2	Problematización .....	9
2.1	Antecedentes .....	9
2.1.1	Educación escolar en Chile .....	9
2.1.2	Educación terciaria en Chile .....	12
2.1.3	Admisión universitaria .....	15
2.1.4	Financiamiento de la educación terciaria .....	19
2.1.5	Brechas de género en la educación terciaria .....	21
2.2	Definición del problema.....	25
2.2.1	Relevancia.....	27
3	Pregunta y objetivos de investigación .....	28
3.1	Objetivos de investigación .....	28
3.1.1	Objetivos específicos.....	28
4	Marco conceptual .....	29
4.1	Trayectorias educativas.....	29
4.2	Género y desigualdad educativa.....	33
5	Marco metodológico.....	38
5.1	Reconstrucción de trayectorias.....	39
5.1.1	Tablas de trayectorias escolares .....	40

5.1.2	Tablas de estado escolar .....	41
5.1.3	Tabla de observaciones escolares.....	42
5.1.4	Codificación individual de programas de estudios en nivel superior.....	42
5.1.5	Tabla de trayectorias terciarias.....	43
5.2	Alcance muestral .....	46
5.3	Procedimientos de análisis exploratorio guiado.....	46
5.3.1	Hitos identificados.....	50
5.4	Declaración de descriptores .....	53
6	Resultados.....	55
6.1	Panorama general de trayectorias en educación superior por género .....	55
6.2	Egreso escolar .....	58
6.3	Postulación al Sistema Único de Admisión .....	66
6.4	Matrícula .....	75
6.5	Retención.....	82
6.6	Reingreso.....	83
7	Discusión .....	88
7.1	Limitaciones .....	91
8	Conclusiones.....	92
	Referencias.....	94
	Tablas.....	101

Anexos .....	109
Anexo A. Identificación del arte mediante una red de referencias .....	109
Anexo B. Estimación de nivel socioeconómico a nivel individual .....	113
Anexo C. Desempeño por modelo-hito .....	115

# 1 Introducción

En la presente investigación se ambiciona indagar en la desigualdad de género en las trayectorias educativas en educación superior en Chile. El paso por la educación terciaria es un fenómeno complejo y relevante en la formación de jóvenes y futuros profesionales que no posee una forma estándar y adquiere múltiples formas de acuerdo con la situación específica de cada estudiante (Blanco et al., 2018).

Algunos aspectos como el género, el nivel socioeconómico, la institución de educación secundaria desde dónde se egresó, el lugar dónde se vive y dónde se estudiará, entre otras múltiples variables configuran diferentes trayectorias educativas (Blanco et al., 2018), limitando o abriendo posibilidades en el desarrollo educativo de jóvenes y adultos que atraviesan la educación terciaria en Chile.

En esto, la existencia de brechas de desigualdad merma tanto el desarrollo individual de los estudiantes como también el desarrollo social a nivel país, limitando la movilidad social, fomentando la reproducción heterogénea de ciertos campos profesionales y cargos laborales y obstaculizando la diversificación de ciertas áreas del conocimiento (Castillo et al., 2014).

El género es uno de los principales aspectos sobre el que se producen una brecha de desigualdad en el ingreso y la trayectoria educativa por el nivel terciario (Bordón et al., 2020; Diaz et al., 2017; Sepúlveda & Manquepillán, 2017), existiendo una vasta literatura que explora el modo en que esta se expresa sobre diversas áreas educativas.

La revisión de esta diversa literatura para el caso chileno (Bordón et al., 2020; Castillo et al., 2014) revela tendencias en ciertas áreas en que esta brecha se muestra más patente: específicamente con la masculinización de las ciencias, la tecnología, las ingenierías y las matemáticas, siendo un fenómeno global y ampliamente estudiado (Ganley et al., 2018; Hägglund & Leuze, 2021; Huang et al., 2020).

Sin embargo, a nivel nacional, existe escasa revisión detallada y en profundidad respecto a este fenómeno específico, que permita describir en detalle cómo el fenómeno de la brecha de género en la educación terciaria se ha expresado ni cómo ha evolucionado junto al aumento de la



matrícula en la educación superior, específicamente en cuanto a la trayectoria de los estudiantes, existiendo revisiones fragmentadas del fenómeno, considerando alguno de los elementos que lo componen pero, por limitaciones técnicas o teorías, no amplían la definición de la trayectoria educativa, siendo el avance más sustantivo lo expuesto por Blanco et al. (2018) y que constituye los cimientos a partir de los cuales profundiza la presente investigación.

Lo que se propone realizar a continuación es una revisión cuantitativa detallada de las trayectorias educativas en el nivel terciario, considerando una conceptualización amplia de estas y que considere las múltiples aristas de que las componen, haciendo particular énfasis en una definición de trayectoria que sea capaz de incorporar a su vez el reingreso a la educación como también la transición entre diferentes programas y áreas de estudio, a fin de identificar, mediante las técnicas de la minería de datos, desigualdades de género que puedan existir en ellas.

## 2 Problematización

### 2.1 Antecedentes

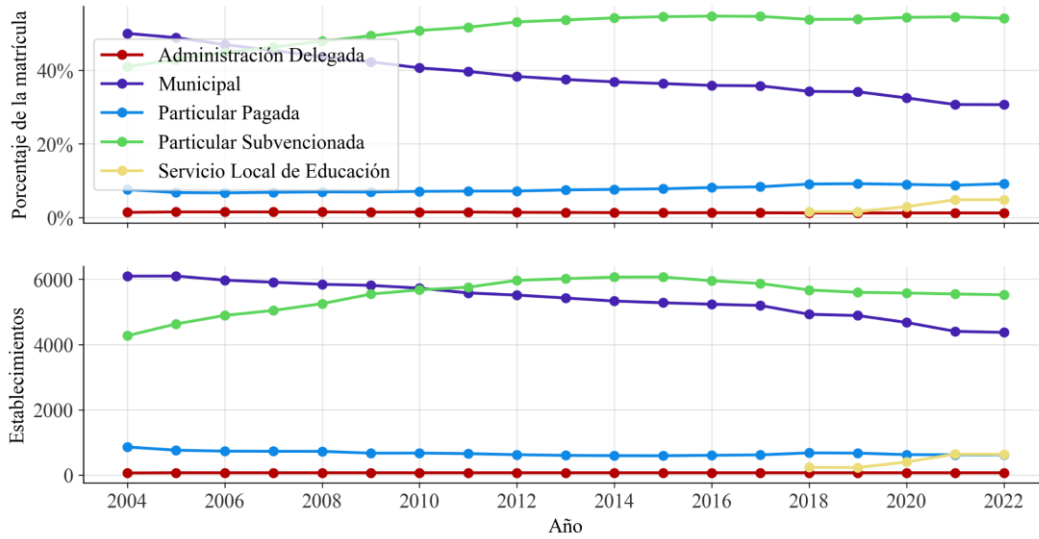
Con el fin de adentrarnos en el problema de investigación, se expone información respecto al contexto educativo en el nivel terciario chileno junto a antecedentes sobre brechas de género en las trayectorias escolares y de educación terciaria.

#### 2.1.1 Educación escolar en Chile

Durante las décadas de 1970 y 1980 se formaron los pilares fundamentales que configuran el sistema educativo escolar chileno mediante la implementación de dos grandes reformas (Venegas-Muggli, 2022). Primero, un nuevo modelo de administración basado en un proceso de descentralización, de modo que los establecimientos escolares dejaron de ser administrados por el Ministerio de Educación para ser transferidos al control de las municipalidades y, segundo, un nuevo modelo de financiamiento, basado en un mecanismo de *vouchers*, así, las escuelas reciben apoyo en relación a la cantidad de estudiantes matriculados, surgiendo con ello establecimientos de dependencia privada con subvención estatal en paralelo a los establecimientos de administración municipal y los de dependencia privada sin financiamiento público (Venegas-Muggli, 2022).

Esta forma de administración sobre los establecimientos educacionales no recibió cambios significativos hasta el período 2014-2018, etapa en que comenzó la implementación (aún en proceso) de algunas modificaciones para, entre otras cosas, prohibir el lucro y la selección en establecimientos que reciben financiamiento público y, además, la instauración de nuevas instituciones públicas (Servicios Locales de Educación) que absorberán la administración de los establecimientos actualmente administrados por las municipalidades (Venegas-Muggli, 2022; Yáñez et al., 2019). Además de lo anterior, existen también los establecimientos con dependencia de Administración Delegada. Estos son 70 liceos de propiedad pública, administrados por 18 corporaciones que proveen educación secundaria técnico-profesional y son financiados públicamente a través de un mecanismo de *charters* (Browne, 2017). La evolución de cantidad de establecimientos y porcentaje de matrícula anual se exponen en la Figura 1.

**Figura 1. Evolución de la cantidad de establecimientos y porcentaje de matrícula por año**

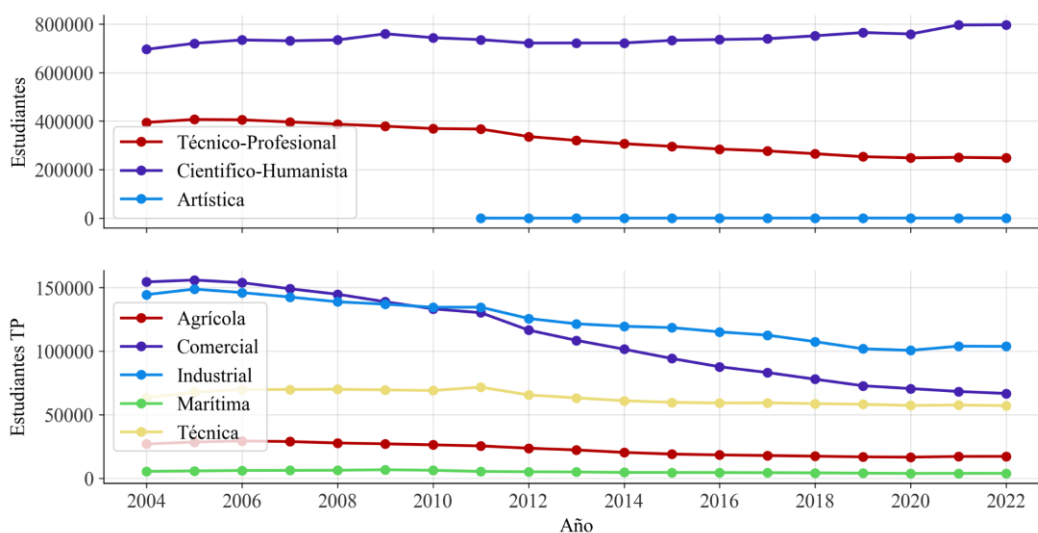


Fuente: Elaboración propia

Sobre esto último, es relevante mencionar que existen tres tipos de educación secundaria, en primer lugar, la educación Científico-Humanista (CH), compuesta por dos opciones electivas (científico o humanista)<sup>1</sup>, luego, la educación Técnico-Profesional (TP), compuesta por ocho ramas de especialidad, y finalmente, la educación Artística, mucho más reducida en su porcentaje de matrícula. La evolución de ello se expone en la Figura 2.

<sup>1</sup> No se dispone de información para conocer la elección de cada estudiante.

**Figura 2. Evolución de estudiantes por tipo de enseñanza media y rama técnico profesional**

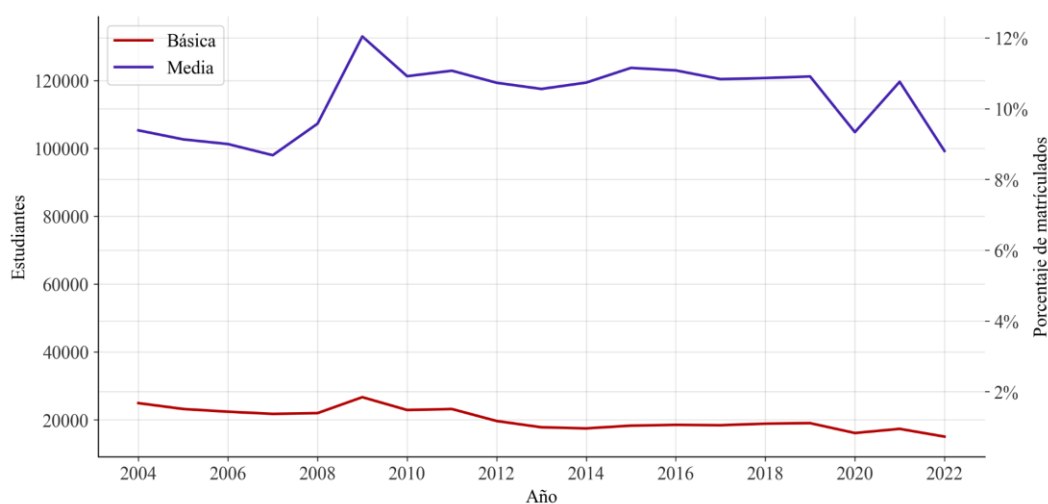


Fuente: Elaboración propia

En el caso de los estudiantes que abandonan su trayectoria escolar y, luego, pretenden regular sus estudios para completar los ciclos primarios y secundarios, obligatorios desde 2003, disponen de la modalidad de educación para jóvenes y adultos (PJA), que consiste en niveles de estudio específico y que permiten, además, la realización de la educación primaria en solo tres años y la educación secundaria en dos (Castillo-Peña, 2021; Diaz et al., 2017). En paralelo, disponen de otros dos mecanismos remediales, estos son una forma flexible, también perteneciente a la modalidad PJA, semi-presencial y dirigida también a jóvenes adultos, orientada a estudiantes que en paralelo trabajan y, finalmente, los proyectos de reinserción escolar, que están orientados específicamente a estudiantes menores de 14 años, financiada por el Ministerio de Educación a través de un fondo concursable al que postulan instituciones públicas y privadas para proveer oferta orientada a la reinserción de menores que hayan desertado previamente del sistema (Diaz et al., 2017)<sup>2</sup>. La evolución de estudiantes en esta modalidad se observa en la Figura 3.

<sup>2</sup> En estos casos, solo se dispone de información sobre el egreso de los estudiantes.

**Figura 3. Estudiantes en modalidad PJA por ciclo educativo (básica o media)**



Fuente: Elaboración propia

### 2.1.2 Educación terciaria en Chile

El sistema de educación terciaria se estructura en tipos de instituciones - Centros de Formación Técnica (CFT), Institutos Profesionales (IP) y universidades- y en niveles cerrados (técnico, profesional y profesional con licenciatura) orientados por fines bien definidos y jerarquizados, considerados cada uno de ellos como niveles terminales y cuya arquitectura supone “un principio de diferenciación jerárquica que sitúa a las universidades en la cúspide del sistema, facultándolas para ofrecer cualquiera de los tres tipos de programas de pregrado” (Gaete & Morales, 2011, p. 52). Mientras que los IP pueden impartir carreras profesionales sin licenciatura y programas técnicos de nivel superior y los CFT solo carreras técnicas.

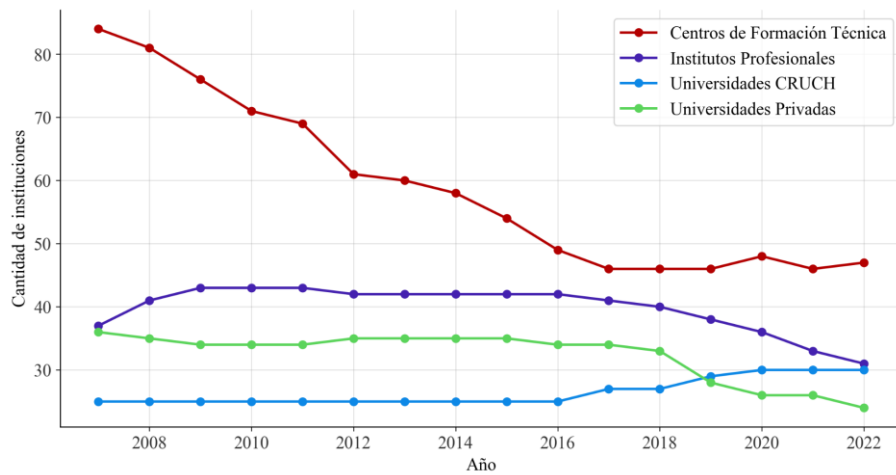
Además, dado los diferentes requisitos económicos implícitos en los costos de estudiar en diferentes instituciones (siendo la educación técnica más accesible que la universitaria), producto de una regulación dada por el mercado, se produce "un sistema de educación terciaria demasiado delimitado, en el que es muy difícil transitar de una modalidad formativa a otra por su falta de flexibilidad" (Gaete & Morales, 2011, p. 54). Solo desde los años 90 es posible hablar en propiedad de un sistema de educación superior, pues hasta los años 80 la educación terciaria se limitaba a un reducido número de instituciones con una oferta restringida al que accedía sólo una élite (Gaete & Morales, 2011).

Junto a esto, "para el caso chileno es importante tener en cuenta que desde el año 1930 se puso en práctica, con carácter obligatorio, un sistema único de selección y admisión de estudiantes a las universidades, basado en un conjunto de pruebas (que han variado sustancialmente en diversos aspectos), que operó como monopolio hasta 1981, cuando las nuevas universidades privadas asumen procesos alternativos de selección. Sin embargo, las universidades que reciben financiamiento público como aportes no vinculantes siguen adscritas a este sistema de selección. Se entiende que este proceso busca, entre otros objetivos, elegir a quienes reúnen las mejores condiciones (de cada promoción) para continuar estudios superiores y, por ende, debería impactar positivamente en la retención estudiantil" (Donoso & Schiefelbein, 2007).

Destaca aquí la distinción entre instituciones selectivas y aquellas que no, considerando que aquellas que forman parte del Consejo de Rectores de las Universidades Chilenas (CRUCH) rigen sus mecanismos de selección por medio de la Prueba de Selección Universitaria (PSU), con un proceso unificado de selección y matrícula de alumnos del cual no forman parte los Centros de Formación Técnica, Institutos Profesionales ni Universidades Privadas no adscritas al CRUCH. De acuerdo con Ramos (2018, p. 20), en esta distinción se materializan nuevas inequidades en la educación superior, pues "la expansión del sistema se produce en instituciones de menor selectividad, [...] la masificación de la cobertura de educación superior viene entonces acompañada de un alto nivel de segmentación social".

Desde finales del siglo anterior, producto del crecimiento del mercado en la educación, hubo un aumento sostenido en la cantidad de instituciones educativas, a la vez que la oferta de matrícula. Según datos del Ministerio de Educación (MINEDUC), en 1980 existían 8 instituciones de educación superior (IES) –todas universidades–. Para el 2010 esa cifra aumentó a 173 instituciones, de las cuales 43 eran institutos profesionales (en adelante IP), 71 centros de formación técnica (en adelante CFT) y 59 universidades, y para 2022, con un notorio descenso en la cantidad de instituciones técnico-profesionales, dichas cifras serían de 31, 47 y 54 respectivamente, totalizando 132 IES.

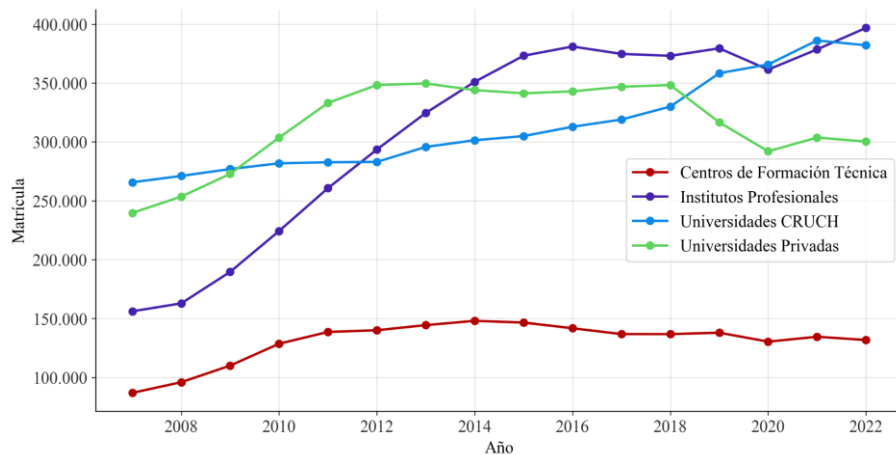
**Figura 4. Evolución de cantidad de instituciones por tipo**



Fuente: Elaboración propia

A pesar del gran número de IP y CFT, la matrícula se concentra en el nivel universitario. Según datos del Servicio de Información de Educación Superior (SIES) del MINEDUC, al año 2022, el 56,3% de la matrícula se concentró en las universidades; el 32,7% en IP y el 10,9% en CFT. Esto muestra un aumento de la oferta de programas de formación de nivel no universitario en instituciones privadas, pero menor matrícula en relación con la demanda total de los estudios universitarios. Además, en la Figura 5 se observa una disminución en la matrícula en universidades privadas respecto a las universidades pertenecientes al CRUCH que, similar a los Institutos Profesionales, han tenido un aumento sostenido de su matrícula en los últimos 15 años.

**Figura 5. Evolución de la matrícula por tipo de institución**



Fuente: Elaboración propia

El incremento exponencial de instituciones universitarias acontecido desde 1981 generó un explosivo crecimiento en las plazas para estudiantes. En los inicios de este proceso de apertura del mercado universitario había cerca de 118.000 estudiantes en las universidades, al año 2005 la cifra se eleva sobre los 450.000, cuya representatividad poblacional saltó del 15% al 37% del grupo en edad de cursar educación superior (Donoso & Schiefelbein, 2007). El crecimiento del número de estudiantes de pregrado del sistema universitario chileno se sustenta – principalmente– en las universidades privadas nacidas a partir de la reforma de 1981, más un aporte significativo de las universidades del Consejo de Rectores (CRUCH).

Existe una crítica a esta organización del sistema educativo de nivel terciario, debido a tres aspectos (Gaete & Morales, 2011, p. 55):

- a) “Un concepto de licenciatura obsoleto que genera barreras artificiales a la movilidad vertical para aquellos jóvenes que, habiendo cursado estudios técnicos o profesionales, desean acceder a programas de pre o posgrado dictados por universidades.
- b) [Oferta] de programas basados en el estatuto legal de las instituciones de educación superior y no en capacidades institucionales demostradas para ofrecer programas de formación técnica o profesional.
- c) Falta de mecanismos y canales que faciliten la transferencia de alumnos dentro y entre instituciones de un mismo sector y nivel de formación profesional y entre instituciones de diferentes sectores y niveles”.

### **2.1.3 Admisión universitaria**

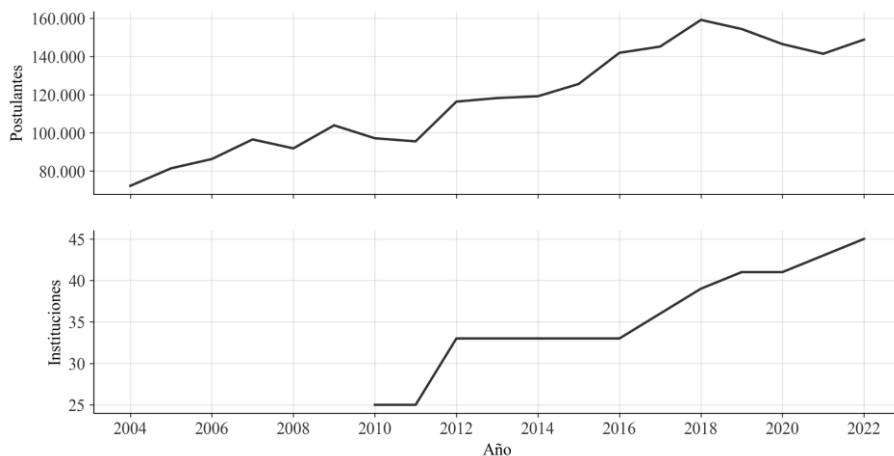
En el caso de las instituciones selectivas, como ya se mencionó, participan de un sistema centralizado de postulación y admisión llamado Sistema Único de Admisión (SUA), administrado originalmente por el Departamento de Evaluación, Medición y Registro



Educacional (DEMRE) de la Universidad de Chile, que está bajo el alero del Consejo de Rectores de las Universidades Chilenas (CRUCH)<sup>3</sup>.

Desde 2003, las universidades pertenecientes al CRUCH han empleado un conjunto de evaluaciones estandarizadas que conforman la Prueba de Selección Universitaria (PSU) para clasificar, ordenar y seleccionar estudiantes. La instalación de un sistema de selección de estudiantes responde a la hipótesis de que no todos los egresados de la enseñanza secundaria poseen las habilidades, condiciones, capacidades, aptitudes o competencias para proseguir exitosamente estudios universitarios. Debiendo someterse a un proceso selectivo cuya razón no se fundamenta en la provisión de más postulantes que vacantes, sino que, en la certificación de un nivel mínimo de desempeño, para augurar buenos resultados en los estudios universitarios, es decir, aumentar la posibilidad de su retención (Donoso & Schiefelbein, 2007). En conjunto con el aumento y expansión de la oferta universitaria junto con la adición de instituciones al SUA, ha aumentado también el número de postulantes, duplicándose en los 18 años con que se encuentra registro, tal como se observa en la Figura 6.

**Figura 6. Evolución de cantidad de postulantes e instituciones en el SUA**



Fuente: Elaboración propia

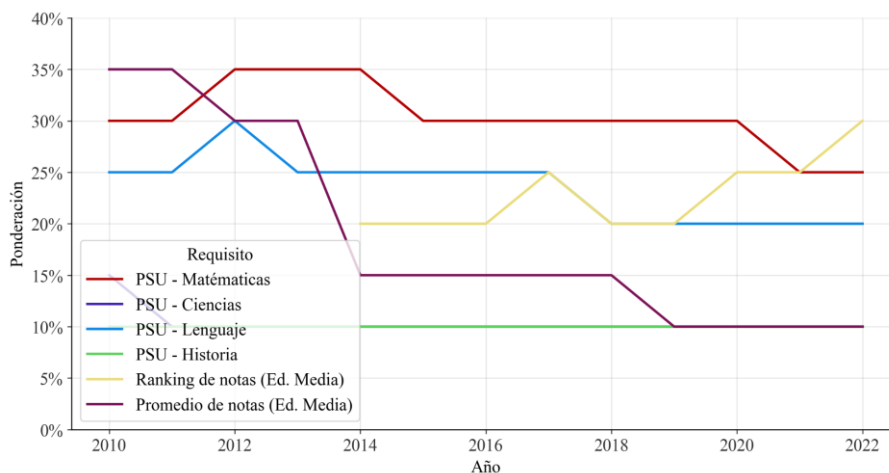
---

<sup>3</sup> Instituciones universitarias tradicionales que reciben financiamiento parcial del Estado.

Todo estudiante que pretenda ingresar a alguna universidad perteneciente al CRUCH debe rendir las evaluaciones de la PSU, tanto las obligatorias (lenguaje y matemáticas) como también al menos una de las opcionales (ciencias sociales y ciencia). Originalmente, el puntaje obtenido era escalado a una distribución normal en el rango de entre 150 y 850 puntos con una mediana de 500. De acuerdo con la literatura relacionada, los resultados en la PSU se correlacionan positivamente con el nivel socioeconómico del estudiante (Pareja et al., 2021).

Una vez recibidos sus puntajes para cada prueba específica, cada estudiante postula a los programas educativos ofrecidos por cada institución participante del SUA, las que publicaban tanto su oferta de vacantes como también los requisitos y ponderaciones específicas de cada una de las pruebas. En función de ello, cada postulación recibe un puntaje asignado en función del desempeño del estudiante en cada prueba y la ponderación de cada una según definió previamente la institución que lo imparte. Destaca en ello el relativo descenso en la ponderación de las pruebas específicas y el aumento de la importancia del ranking de notas de enseñanza media, como se observa en la Figura 7.

**Figura 7. Evolución de la mediana de ponderación por requisito de ingreso en el SUA**

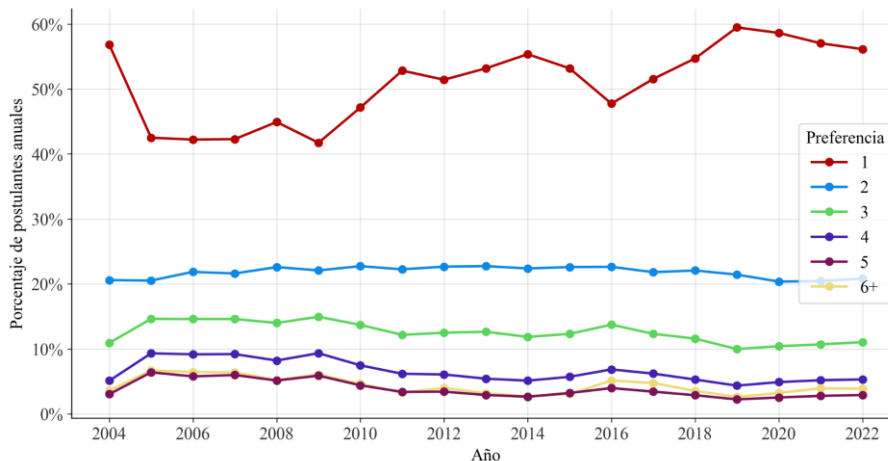


Fuente: Elaboración propia

Así el sistema completa sus vacantes en un orden descendente y, si un estudiante es seleccionado en su primera opción, será descartado del resto de postulaciones, hasta iterativamente completarse las vacantes de cada programa ofrecido. Es destacable el nivel de eficiencia de dicho mecanismo, de manera tal que entre un 40% y un 60% de los estudiantes es asignado en su primera preferencia, cifra que ronda el 70% si se considera la primera y segunda

preferencia, con menos de un 10% de los postulantes siendo asignados a una carrera posterior a su quinta preferencia, tal y como se visualiza en la Figura 8.

**Figura 8. Evolución del porcentaje de seleccionados por preferencia en el SUA**



Fuente: Elaboración propia

Adicionalmente, las instituciones publican los puntajes de corte (el puntaje del último matriculado) del proceso de admisión anterior, de modo que cada postulante cuenta con información completa sobre el puntaje con que postula a un programa y su distancia al último matriculado en el proceso anterior, por lo que tiene incentivo para ordenar sus postulaciones según sus preferencias para priorizar un buen resultado en su admisión.

Se debe comentar además que, a partir de 2020, comenzó una transición para migrar la gestión sobre el sistema de admisión desde el DEMRE hacia la Subsecretaría de Educación Superior (SES) dependiente del Ministerio de Educación, para lo que se implementó de manera transitoria una prueba denominada Prueba de Transición Universitaria (PDT/PTU), con diferencias curriculares en la elaboración y contenido a evaluar por cada una de las pruebas.

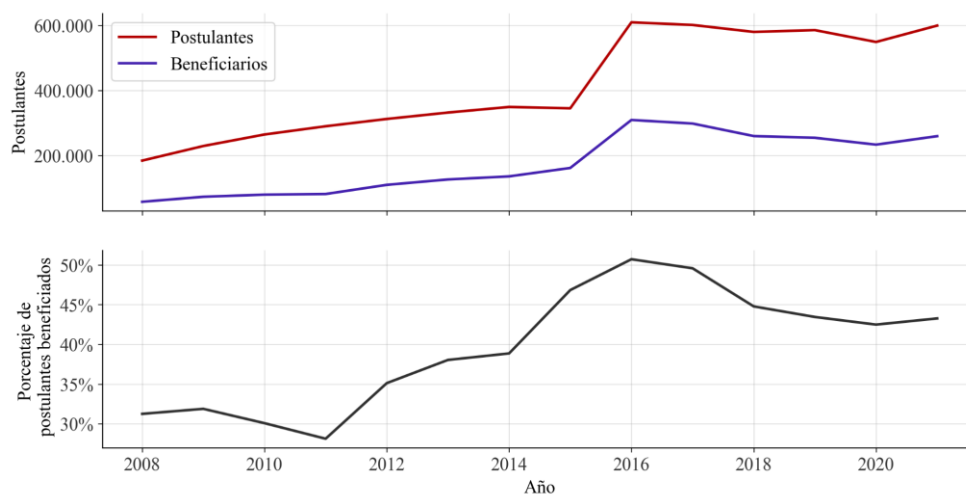
Lo anterior culminó en 2022 con la presentación de la Prueba de Acceso a la Educación Superior (PAES) que presenta diferencias respecto a la PSU, principalmente en la escala de puntajes, al no estandarizar la distribución de resultados y preservarlos en un rango de 100 a 1.000 puntos, ello con el objetivo de facilitar la comparabilidad de puntajes entre procesos, además de otros aspectos como una mayor orientación a la evaluación de habilidades en desmedro a la evaluación de conocimientos.

## 2.1.4 Financiamiento de la educación terciaria

La educación superior chilena, desde la perspectiva del estudiante y la demanda, es financiada principalmente mediante cuatro vías (CPCE, 2020):

- a) El pago directo desde el estudiante y su familia.
- b) Las ayudas institucionales como becas internas de cada institución de educación superior.
- c) Las ayudas estatales como becas de educación superior provistas por el Estado<sup>4</sup>.
- d) El endeudamiento en la forma de créditos estudiantiles, destacando principalmente el Crédito con Aval del Estado o Crédito de Garantía Estatal (CAE) y el Fondo Solidario de Crédito Universitario (FSCU).

**Figura 9. Evolución de postulantes y beneficiarios a beneficios estatales**



Fuente: Elaboración propia

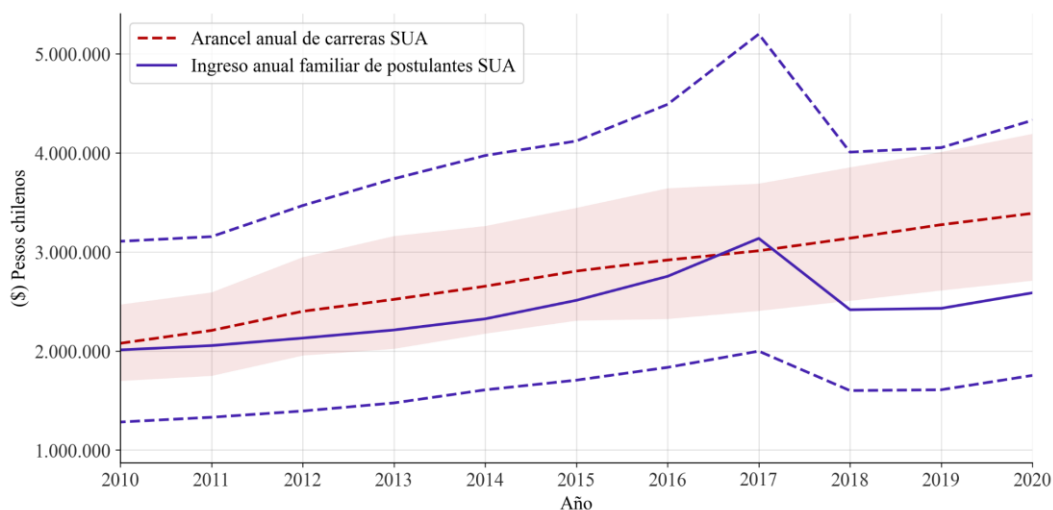
En la Figura 9 se visualiza la evolución de los postulantes y receptores de los beneficios antes listados.

---

<sup>4</sup> También existen las becas de mantención, principalmente aquellas financiadas por la Junta Nacional de Auxilio y Becas (JUNAEB), pese a que ellas no son consideradas en la presente investigación por la escasa información sistematizada disponible.

La posibilidad de pago directo del estudiante y su familia es limitada por el alto valor de los aranceles de las diferentes carreras, tal y como se observa en la Figura 10, siendo necesario, para expandir el acceso a la educación superior hacia sectores de menores ingresos, la existencia de las otras tres maneras comentadas anteriormente. Ello considerando que, en general, los jóvenes de sectores populares estudian menos años y entran a trabajar a edades más tempranas (Pareja et al., 2021).

**Figura 10. Evolución del arancel anual y del ingreso familiar de los postulantes a programas en el SUA<sup>5</sup>**

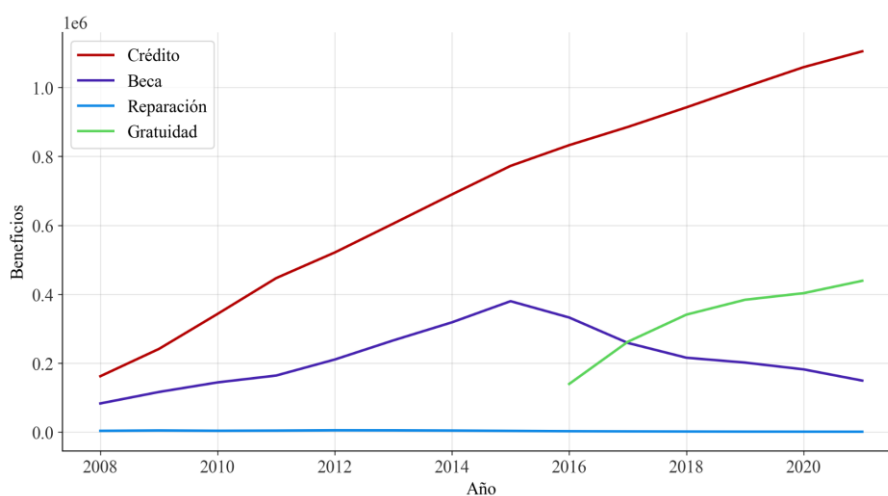


Fuente: Elaboración propia

De estos, el endeudamiento es la principal fuente de financiamiento, con un aumento sostenido en la cantidad de créditos otorgados desde que se disponen registros accesibles, sin embargo, destaca que la gratuidad, desde su implementación, ha reemplazado fuertemente a las otras formas de becas sobre el arancel, lo que es positivo para reducir desigualdades socioeconómicas de acuerdo con Pareja et al. (2021), quien observa que el acceso al crédito mitiga notoriamente las diferencias en matrícula por estrato socioeconómico.

<sup>5</sup> El área roja representa el segmento entre el tercer y el primer cuartil de aranceles, mientras las líneas azules visualizan dichos cuartiles para el ingreso familiar

**Figura 11. Evolución de beneficios asignados por tipo**



Fuente: Elaboración propia

En paralelo, sin embargo, destaca en la Figura 11 que la gratuidad, desde su implementación, ha reemplazado fuertemente a las otras formas de becas sobre el arancel. A diferencia de las becas de arancel tradicionales, es un mecanismo de financiamiento que no considera requisitos académicos (Delisle & Bernasconi, 2018), aunque no todas las instituciones son elegibles para disponer de dicho beneficio.

### **2.1.5 Brechas de género en la educación terciaria**

La existencia de brechas de género en diferentes áreas y disciplinas es un asunto conocido, existiendo diferentes organizaciones e instituciones que realizan una labor destacable con la intención de diversificar áreas del conocimiento en que el género femenino ha sido, según sus diagnósticos, históricamente excluido.

De acuerdo con Carrasco (2016), la brecha más visible en cuanto a acceso y permanencia en la Educación Superior es socioeconómica, pues son los jóvenes de sectores más desfavorecidos desde este punto de vista los que más han visto perjudicada su capacidad de elegir carrera libremente.

De igual modo, si bien la brecha socioeconómica es profundamente relevante en el acceso y permanencia en la educación terciaria chilena, hay que reconocer la existencia de otras brechas (Sepúlveda & Manquepillán, 2017), como la pertenencia a pueblos originarios,

ruralidad, discapacidad y género, entendiendo este último como la separación entre lo masculino y lo femenino al interior de la educación superior.

A nivel global y regional, en el ámbito académico y profesional, este fenómeno de distinción de lo masculino y lo femenino se expresa en situaciones tales como la subrepresentación femenina en disciplinas asociadas a las ciencias, tecnología, ingeniería y matemáticas (Castillo et al., 2014), lo que parece estar asociado a una serie de obstáculos a la participación de mujeres en estas áreas, como son las preferencias, motivos, valores, estereotipos y normas culturales asociadas a una distinción de lo femenino y lo masculino que merma la participación de las mujeres.

Este fenómeno produce la rotulación de carreras como "masculinas" o "femeninas" sustentada en "prejuicios, juicios de valor, marcos de referencia, estereotipos y representaciones sociales arraigados al interior de la sociedad" (Peña y Lillo, 2006).

Existe multitud de explicaciones que buscan determinar las variables fundamentales de esta desigualdad, con aproximaciones biológicas, sociales, culturales y hasta institucionales y, si bien no hay un consenso al respecto, se observa que a edad temprana no existen elementos de desventaja para las niñas por sobre los niños en cuanto a sus habilidades cognitivas y motoras, de modo que las diferencias que se presenten posteriormente obedecerán a variables contextuales y de socialización que estarían permeado fuertemente el sistema educativo (Sepúlveda & Manquepillán, 2017).

Esto deja en evidencia la existencia de procesos de socialización realizados desde diversas instituciones y agentes que producen y/o reproducen discursos que definen diversidad masculinidades o feminidades (Peña y Lillo, 2006; Sepúlveda & Manquepillán, 2017).

Esta situación se relaciona junto a otros factores de inequidad en la educación, en un entramado complejo de factores que inciden sobre el tránsito de los estudiantes por las carreras educativas en el nivel superior, generando así una pluralidad de trayectorias educativas y transiciones.

Al respecto, para las estudiantes mujeres, de acuerdo con Guevara y García (2010), implica una serie de presiones en la vida estudiantil y académica, tales como la visión de ciertas

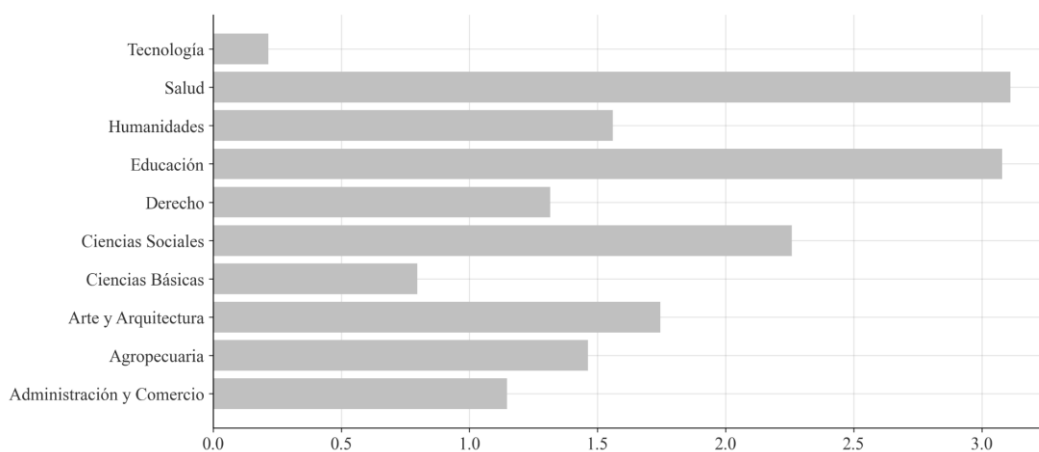
áreas como un entorno en que no pertenecen, una mayor exigencia por cumplir con labores domésticas, la percepción de ciertas profesiones como masculinas y, por tanto, impropias para ellas, entre otros.

Estas situaciones conllevan una trayectoria académica potencialmente asimétrica entre hombres y mujeres que, en áreas específicas, como las ciencias y la tecnología se ven incrementadas (Castillo et al., 2014; Guevara & García, 2010; Hägglund & Leuze, 2021; Sepúlveda & Manquepillán, 2017).

Esta situación no es solo de asignación de géneros, sino también adecuada a una percepción sobre las carreras, en tanto carreras identificadas como masculinas tienen un accionar que es sobredimensionado, mientras que aquellas que son identificadas como femeninas se ven disminuidas (Baeza & Lamadrid, 2018), situación que profundiza la brecha de género en la matrícula de estas carreras.

Como se observa en el siguiente gráfico, existe una brecha de género visible en la matrícula por área del conocimiento, en que tecnología aparece marcadamente masculina, mientras las áreas de educación y salud son predominantemente femeninas.

**Figura 12. Proporción de mujeres respecto a hombres por área del conocimiento**



Fuente: Elaboración propia

Esto permite vislumbrar la existencia a nivel nacional de carreras feminizadas y masculinizadas: "Desde sus orígenes, la ingeniería ha estado típicamente liderada por hombres.



[...] En tanto, el trabajo social desde sus inicios se modeló como una carrera netamente femenina" (Peña y Lillo, 2006).

Así, se observa cómo la educación superior en Chile se configura con un carácter desigual, en que la participación al interior de los programas de estudio está atravesada por brechas de desigualdad entre grupos, de los cuales destacan aquellos definidos por su situación socioeconómica y por su género. A nivel general, Chile es un país particularmente afectado por brechas de género en detrimento de las mujeres, ocupando el lugar 70 de 144 naciones en el Índice de Brechas de Género del Foro Económico Mundial (Obregón & Sepúlveda, 2019).

Es pertinente mencionar que no es posible ignorar la existencia de múltiples iniciativas que justamente responden al diagnóstico de este problema, de las cuales es posible comentar algunos ejemplos:

- Laboratoria<sup>6</sup> realiza una doble labor en la formación e inserción laboral de mujeres en el área de la computación, específicamente en el desarrollo frontend y en el diseño UX, siguiendo el objetivo de diversificar el campo del desarrollo web y nutrir una carencia de profesionales en esta área.
- NiñasPRO<sup>7</sup> sigue una labor similar, fomentando el interés y el ingreso de niñas y adolescentes en el campo de la informática, realizando diversos talleres de programación para este sector demográfico.
- Ingeniosas<sup>8</sup> es una iniciativa para niñas que busca acercar a ellas una perspectiva sobre la tecnología, el arte, la ingeniería y la matemática que discuta los estereotipos y prejuicios existentes en estas áreas, para así fomentar la vocación de las jóvenes hacia estas áreas.

---

<sup>6</sup> Ver más en: <https://www.laboratoria.la/>

<sup>7</sup> Más información en: <https://ninaspro.cl/somos/>

<sup>8</sup> Detalle en: <https://ingeniosas.org/quienes-somos/>

- GirlsInTech<sup>9</sup> es una organización sin fines de lucro que se dedica a lidiar con la brecha de género en tecnología mediante la realización de bootcamps, apoyo vocacional, conferencias, entre otros.
- La Universidad de Chile<sup>10</sup>, específicamente la Facultad de Ciencias Físicas y Matemáticas, privilegia mediante cupos de ingreso y prioridad en las listas de espera el ingreso femenino.
- La Universidad Austral de Chile y su facultad de Ciencias de la Ingeniería<sup>11</sup> cuenta con el programa "Más Mujeres en Ingeniería", mediante el cual busca fomentar el ingreso y la permanencia de mujeres en carreras de Ingeniería mediante cupos de admisión especial.

Estos casos ejemplifican el reconocimiento de la trascendencia del género como un eje articulador de la trayectoria educativa de las estudiantes y cómo su marcada relevancia en determinadas áreas ha marcado un camino para orientar las intervenciones que buscan un mejoramiento en las brechas de desigualdad educativa en la educación terciaria en Chile. Tras todos estos programas se halla la premisa fundamental de que el género (y más específicamente, ser mujer) tiene implicancias sobre la trayectoria en la educación terciaria, limitando las posibilidades de ingreso y permanencia en los estudios, lo que es aún más marcado en determinadas áreas.

## 2.2 Definición del problema

El problema sobre el que se pretende investigar se inserta en el fenómeno global de las desigualdades de género en la educación. Como ya se comentó, esta situación se hace visible en la educación terciaria chilena al observar la composición por género de ciertas carreras, sin

---

<sup>9</sup> Más detalles en: <https://chile.girlsintech.org/about/>

<sup>10</sup> Como se detalla en: <http://ingenieria.uchile.cl/admision/admision-especial-pregrado/94355/cupos-equidad-de-genero>

<sup>11</sup> Según se informa en: <https://diario.uach.cl/mas-mujeres-en-ingenieria-uach-pone-en-marcha-nuevo-programa-de-ingreso-especial/>

embargo, una cuestión pendiente corresponde al carácter general de este fenómeno, es decir, cuantificar la situación de las mujeres en la educación terciaria en el tiempo reciente a fin de identificar el modo específico y detallado en que se manifiestan las asimetrías descritas.

En función de los antecedentes anteriormente expuestos, se afirma que, primero, existe la desigualdad de género y, segundo, que esta desigualdad se manifiesta en la educación superior.

Al respecto, los antecedentes empíricos que buscan explicar los fenómenos asociados a las trayectorias educativas, principalmente en cuanto a abandono y reingreso, desde su definición más tradicional han considerado una serie de dimensiones y aspectos asociados a los estudiantes, su entorno, el contexto institucional, entre otros.

Sin embargo, pese a la multitud de evidencia empírica que sustenta la relevancia del género como un aspecto clave en la producción y reproducción de desigualdades en lo social y lo educativo, la evaluación específica de este en las trayectorias educativas ha ocupado una posición limitada, siendo incipientes las investigaciones que intentan dar cuenta de las diferencias existentes en torno a esto, a la vez que la consideración sobre el fenómeno de las trayectorias educativas en educación superior se abarca desde aproximaciones fragmentadas, que por diferentes limitaciones o decisiones, consideran versiones simplificadas o reducidas del concepto, realizando un seguimiento limitado de los estudiantes.

A raíz de lo expuesto, se propone investigar sobre las desigualdades de género en las trayectorias en la educación superior chilena entre los años 2008 y 2021, desde una definición de trayectoria que permita considerar una mayor complejidad y detalle respecto a los procesos que las componen, de este modo se propone participar de dos asuntos que serán profundizados posteriormente: Primero, la necesidad de desarrollar la noción de trayectorias educativas en el nivel terciario más allá de la retención y la deserción, siendo necesario reconocer y detallar la posibilidad de reingreso y tránsito entre diferentes programas de estudio y, segundo, respecto a este tránsito, incorporar los antecedentes sobre las diferencias en la participación masculina y femenina en diferentes áreas del conocimiento, que actualmente han tomado un rol incipiente y aún distante en la investigación sobre trayectorias académicas.

### **2.2.1 Relevancia**

En primer lugar, el problema de la inequidad en la educación en que se conoce la existencia de brechas en los procesos de postulación, ingreso y titulación en los programas de estudio, lo que otorga relevancia social a profundizar en el tema seleccionado, a fin de producir información que sirva de antecedente para conocer y actuar sobre el problema y así hallarle soluciones. Destaca además el hecho de estar frente a un problema cruzado por múltiples causas, globales y locales, que requiere entonces de una mirada en profundidad sobre la forma en que se manifiesta como fenómeno.

Segundo, en relación con lo anterior, si bien existen investigaciones que consideran el problema aquí presentado (Blanco et al., 2018), así como también otros asuntos relacionados, no existen antecedentes que cuantifiquen en profundidad y a nivel nacional la desigualdad de género en trayectorias educativas de nivel terciario desde una definición que abarque el fenómeno completo. Por lo tanto, es relevante ahondar en la relación de esta con asuntos tales como las brechas en cada proceso asociado a las trayectorias, la distinción entre instituciones selectivas y no selectivas, el tránsito entre áreas del conocimiento y educativas masculinizadas o feminizadas, entre otros.

Finalmente, volviendo al problema de la desigualdad en la educación, existiendo una participación desigual de ciertos grupos en determinados programas de estudio, se genera una falta de diversidad en el área laboral, profesional y académica, lo que de acuerdo con la literatura (Castillo et al., 2014) es una restricción a la innovación, mermando el crecimiento productivo y el desarrollo económico, a la par que genera campos productivos con una visión sesgada sobre los problemas en los que actúan y el carácter de las necesidades que buscan satisfacer, además, junto a lo anterior, la evidencia indica que las áreas masculinizadas coinciden en buena parte con aquellas que tienen mayor ingreso (Bordón et al., 2020), por lo que la desigualdad de género incide potencialmente en una desigualdad de ingresos.

### **3 Pregunta y objetivos de investigación**

La pregunta de investigación que se propone responder es: ¿Cómo se expresan cuantitativamente las desigualdades de género en las trayectorias educativas en el nivel terciario de la educación chilena en el periodo 2008-2021?

#### **3.1 Objetivos de investigación**

Objetivo principal: Caracterizar las desigualdades de género en las trayectorias educativas en el nivel terciario de la educación chilena en el periodo 2007-2019.

##### **3.1.1 Objetivos específicos**

- Describir brechas de desigualdad en los diferentes hitos que componen y describen las trayectorias estudiantiles en el nivel terciario.
- Identificar descriptores relevantes en la elección de programas de estudio en relación con el género del estudiante.
- Identificar la existencia de sub-brechas de desigualdad que se produzcan en la interacción del género con diferentes descriptores.

## 4 Marco conceptual

A continuación, se expone una discusión teóricamente fundamentada sobre los supuestos y conceptos que se emplearán en los apartados subsecuentes. Esta discusión se divide en dos etapas, primero, una exposición sobre los antecedentes directos que orientan el uso de trayectorias educativas como artefacto analítico en educación terciaria, fundamentada principalmente en el trabajo de Blanco et al. (2018) y luego una discusión respecto a las implicancias del género como elemento articulador de decisiones en la vida individual de las personas, las causas de dichas diferenciaciones y sus consecuencias.

A diferencia del primer punto, que está basado principalmente en un único trabajo anterior, la discusión sobre el género y la educación es un tópico de investigación muy activo, por lo que la revisión bibliográfica se realizó mediante una red de citas, con un procedimiento descrito en el Anexo A. De ello es destacable que considerando los trabajos publicados desde 2016, únicamente artículos de revistas no retratados y que contuviesen en su título o resumen las palabras género y escuela, universidad, STEM o educación, con una distancia máxima a las semillas<sup>12</sup> de 3 grados, se obtuvo un total de 3.412 trabajos relacionados. De esto, y siguiendo lo ya expuesto, se seleccionaron los siete<sup>13</sup> artículos de mayor relevancia para la revisión del estado del arte del tópico de investigación.

### 4.1 Trayectorias educativas

De acuerdo con Baeza y Lamadrid (2018), la evidencia indica que el género influye desfavorablemente en la educación y, como indican Bordón et al. (2020), pese a los notorios esfuerzos relacionados al cierre de brechas de género en los años de escolaridad y en asistencia a educación secundaria y postsecundaria, su presencia en las diferentes disciplinas es desigual,

---

<sup>12</sup> Considerando como semillas los siguientes artículos (por su DOI): 10.1016/j.econedurev.2020.102011; 10.1080/02601370.2022.2041746; 10.1080/03057925.2019.1675495; 10.1080/03055698.2022.2117540

<sup>13</sup> En un umbral definido de acuerdo con la contribución general de adicionar un nuevo artículo a la revisión respecto a la información ya declarada en aquellos de mayor relevancia.

estando concentrada, en términos internacionales, en las áreas de humanidades, educación, salud y artes, y severamente subrepresentada en las ingenierías y en el área de tecnologías de información y comunicación.

Lo anterior, sumado a como los antecedentes empíricos dan cuenta de que el acceso a la educación superior mejora la noción sobre sí mismas y su participación en el espacio público para las mujeres (Baeza & Lamadrid, 2018), sugiere la necesidad de indagar en las diferentes configuraciones que ocurren en el tránsito hacia y durante la educación superior en función del género de los estudiantes.

De acuerdo con Baeza y Lamadrid (2018), es notorio cómo los estudiantes reconocen las diferencias de género en las carreras profesionales, subrayando aspectos como las aspiraciones diferenciadas, pero también la desigual distribución de género en carreras como ingeniería y educación parvularia, identificándolas según lo “típico” (o lo “adecuado”) y lo “raro”, de acuerdo también a roles diferenciados en la división de labores dentro de la familia, aunque reconociendo también las tendencias de cambio en ello.

De acuerdo con Bordón et al. (2020), al interior del SUA, dichos elementos se materializan en una postulación diferenciada por género en que las áreas de tecnología e ingeniería son una opción más probable para postulantes hombres respecto a las mujeres y, en sentido opuesto, las estudiantes de género femenino están más dispuestas a postular a las áreas de salud. Ello da cuenta de cómo, incluso previo al ingreso a la educación superior, la trayectoria de los estudiantes ya presenta brechas que se gestan desde sus propias intenciones y percepciones respecto a aquellas áreas que les interesa estudiar.

En el campo de las trayectorias educativas en la educación superior chilena, Blanco et al. (2018) hacen una contribución inédita al realizar un análisis a las dinámicas de titulación, permanencia, deserción y reingreso a la educación superior a partir de datos administrativos que permiten trazar el tránsito de un estudiante desde su egreso hasta los resultados de su paso por la educación terciaria (ya sea su titulación o retiro, considerando también posibles tránsitos entre programas de estudios).

Al respecto, los autores exponen la necesidad de complejizar las definiciones de trayectorias más allá del binarismo éxito-fracaso académico que es propio de perspectivas de gestión institucional y políticas públicas que ponen su foco en la retención y, por tanto, clasifican bajo la conceptualización del “fracaso” todo movimiento del estudiante que no incide en su permanencia y eventual titulación del programa de estudios.

En este sentido, y fomentado por la capacidad de acceso a datos administrativos cuasi-censales, se aproximan al fenómeno de las trayectorias con un foco puesto en la posibilidad de diversas trayectorias posibles (y válidas), complejizando el interés sobre el tópico investigativo que permita dar cuenta de otros fenómenos relacionados.

En cuanto a los antecedentes a ello, destaca, como ya se mencionó, el interés sobre la deserción (Blanco et al., 2018), que en general se relaciona a aspectos socioeconómicos, personales, institucionales y académicos, reconociendo la importancia de identificar el momento de la formación en que ello ocurre, pudiendo reconocer etapas críticas como la admisión, los primeros momentos desde el ingreso, el final del primer año o comienzos del segundo, pudiendo clasificarse así como deserción inicial, temprana o tardía y también en función del espacio desde donde se deserta (programa, institución o sistema) (Blanco et al., 2018; Fincham et al., 2021).

Diferentes fuentes y aproximaciones disciplinarias destacan diferentes aspectos para dar cuenta de los orígenes y causas de dicho evento (Fincham et al., 2021), ya sea como las aproximaciones psicológicas que otorgan relevancia al rol de los rasgos de personalidad, actitud individual y expectativas sociales, los modelos sociológicos que dan cuenta de aspectos de integración social, rendimiento académico, satisfacción del estudiante, compromiso institucional y condiciones socioeconómicas y socioculturales o, como los modelos económicos, considerando la deserción como una toma de decisiones relativa al cálculo sobre la ganancia y los costos implicados en la completación de un programa de estudios y del retiro desde este.

A partir de dichas explicaciones, se han elaborado modelos causales integrales que consideran un crisol de dichas explicaciones para elaborar aproximaciones de mayor sofisticación como el *Student Integration Model* (SIM) que propone que “niveles deficitarios de integración estudiantil en las dimensiones académica y social definen las decisiones de los estudiantes de permanecer o abandonar, [...] Las decisiones de deserción se relacionan con la



evolución de elementos tanto del agente como de la estructura institucional, incluyendo los niveles de compromiso con las metas personales y de compromiso institucional. Estos procesos tienen a su vez como entradas las calificaciones previas, los atributos individuales y familiares, así como las condiciones contextuales pedagógicas y sociales.” (Blanco et al., 2018, p. 145) o la *mismatch hypothesis* que “se centra en el ajuste entre el estudiante, la institución y el programa de estudios, considerando tanto factores económicos, psicológicos y sociales. [Así] evalúa si el estudiante y la institución se ajustan y son un buen complemento, analizando la posibilidad de que los estudiantes estuviesen mejor en otra institución o programa” (Blanco et al., 2018, p. 145).

En paralelo, se considera también el fenómeno del *stopout* o reingreso de corto plazo, esto es, aquellos estudiantes que suspenden sus estudios y vuelven a ingresar posteriormente en un corto plazo (que, en antecedentes empíricos se define como un máximo de tres años), sin embargo, los autores puntualizan que las posibilidades de revisar dicho fenómeno requiere, primero, de una forma de estudios con una cierta flexibilidad que permita la suspensión de estudios y, en paralelo, de una aproximación teórica adecuada que vaya más allá de la reducción binaria basada en el éxito y el fracaso académico, pues “el estudiante que no se gradúa ni permanece en su institución y programa de estudio puede desertar definitivamente del sistema, o bien, reingresar a distintas alternativas de educación superior” (Blanco et al., 2018, p. 147).

En sus resultados, Blanco et al. (2018) dan cuenta de que, tras nueve años, un 25% de los estudiantes desertó sin reingreso, mientras que un 60% se ha titulado y un 15% continúa estudiando, revelando, a la vez, la ineficiencia del sistema educativo respecto a los tiempos de estudios y su flexibilidad para captar reingresos y, por tanto, efectivo en la titulación, con un ligero sesgo que favorece a las estudiantes de género femenino en los tiempos de titulación más breves, aunque ello también ocurre respecto a mejores condiciones académicas y socioeconómicas. Además, la mitad de los estudiantes que ingresan transitan hacia un área educativa diferente a la original, dando cuenta de tránsitos potencialmente relacionados a la orientación vocacional.

Sobre la titulación, se detalla que los reingresos tienen una tasa de completación de estudios que ronda el 50% y que es, sin embargo, inferior a la de aquellos que permanecen en sus carreras originales, siendo esta diferencia inferior en el sector universitario que transita hacia la

educación técnico profesional, ello posiblemente relacionado a la longitud de las carreras (Blanco et al., 2018).

Entre instituciones, dichas diferencias presentan una alta variabilidad en ambos sectores, sin embargo, en instituciones técnicas se presenta una mayor titulación y retención que en universidades privadas, lo que los autores hipotetizan estaría vinculado con la duración de las carreras.

## **4.2 Género y desigualdad educativa**

La diferenciación por género en diferentes aspectos de la vida ha tenido notoria relevancia en años recientes (Charlesworth & Banaji, 2019; Moss-Racusin et al., 2018; Stewart-Williams & Halsey, 2021) y, puntualmente, en el campo de lo ocupacional, destaca el debate de la desigual distribución de personas masculinas y femeninas en los diferentes campos del conocimiento y disciplinas (Moss-Racusin et al., 2018; Richardson et al., 2020), siendo ello una discusión muy activa específicamente en el tópico de las áreas STEM (Ciencia, Tecnología, Ingeniería y Matemáticas, por sus siglas en inglés), escenario en que la idea de que la presencia masculina supera a la femenina ha pasado a ser la convención a lo largo de las últimas décadas (Moss-Racusin et al., 2018; Richardson et al., 2020; Stewart-Williams & Halsey, 2021), a pesar de que de acuerdo a ciertos análisis, dicha sobrerrepresentación no es ubicua (Stoet & Geary, 2018; Thébaud & Charles, 2018), aunque ello es una discusión aún existente, considerando cómo Stoet (2018) observa que, paradójicamente, países con mayor igualdad de género institucional también exhiben brechas en la participación femenina en áreas de ciencia y tecnología más marcadas, aunque ello es discutido por Richardson et al. (2020), al notar que es necesario considerar también aspectos ulteriores como la permanencia en los programas y su propensión a graduarse.

Al respecto, un primer elemento conceptual a considerar es que el campo STEM no homogéneo en su interior sobre las disciplinas que abarca, existiendo las diferencias entre las áreas GEEMP (Geociencias, ingeniería, economía, matemáticas/ciencias de la computación y ciencias físicas, en inglés) y las LPS (Ciencias de la vida, psicología y ciencias sociales, también por sus siglas en inglés). Distinción en la que, análisis similares, han dado cuenta de que el género masculino está sobrerrepresentado únicamente en GEEMP, mientras que en LPS existe

igualdad si es que no sobrerrepresentación femenina (Stewart-Williams & Halsey, 2021), a la par que ciertos autores como Cheryan (2017) dan cuenta acerca de cómo no todos los campos clasificados como STEM han sido igualmente estudiados.

Las aproximaciones que buscan explicar dichas brechas en la representación de ambos géneros en estas áreas pasan por, primero, explicaciones biológicas (Stewart), o también denominadas esencialistas (Thébaud & Charles, 2018), que pretenden dar cuenta de las causas naturales del asunto en función de intereses innatos diferenciados, como lo comentado por Stewart (2021) sobre evidencia que respalda un interés diferenciado en el eje “interés sobre las cosas” e “interés sobre las personas”, sobre lo que dicho autor comenta que diferentes estudios realizados en humanos y mamíferos primates exhiben resultados similares en que sean aquellos individuos de sexo femenino quienes presentan mayor posicionamiento en el “interés sobre las personas”, en simetría opuesta a aquellos de sexo masculino, quienes están más dispuestos al “interés sobre las cosas”.

Es el mismo autor quien puntualiza que, además de ello, el sexo masculino otorga mayor prioridad en la vida al éxito y los ingresos, respecto al sexo femenino, cuyos miembros reportan mayor relevancia al balance entre vida y trabajo personal, según se sustenta en un estudio longitudinal realizado sobre una muestra de individuos identificados como intelectualmente aventajados en su adolescencia temprana (Stewart-Williams & Halsey, 2021), situación que también incidiría en una desigual distribución en horas ocupadas hacia trabajos remunerados y, además, en aptitudes cognitivas, sobre lo que Stewart (2021) comenta que, si bien las diferencias halladas en investigaciones previas no son una magnitud notoria, son los miembros del sexo masculino quienes muestran un mejor desempeño en evaluaciones de habilidades espaciales, razonamiento mecánico y habilidades matemáticas, mientras que quienes pertenecen al sexo femenino tendrían mejor desempeño en habilidades del lenguaje como la comprensión verbal, la lectura y la escritura.

Todo ello, de acuerdo con Stewart (2021), respaldado en una racionalidad evolutiva relacionada al tiempo invertido, por condiciones biológicas, a las labores de crianza, sin embargo, autores como Thebaud & Charles (2018), destacan que dichas explicaciones no dan cuenta de las razones por las que existe tanta variación en dichas brechas a lo largo del tiempo y

en diferentes territorios y grupos demográficos (Thébaud & Charles, 2018) y que además interactúa con otros aspectos sociodemográficos como la etnicidad, la situación socioeconómica y el estatus migratorio (Thébaud & Charles, 2018), sugiriendo que ciertos estereotipos sobre de la cultura occidental sobre la naturaleza de las ocupaciones STEM y sus trabajadores, en interacción con “las cualidades intrínsecas y el estatus social relativo a cada sujeto, pueden ser marcados conductores de afinidades y aspiraciones” (Cheryan et al., 2017; Thébaud & Charles, 2018).

Al respecto, además, Thebaud & Charles (2018) hace notar que, indiferente de potenciales explicaciones basadas en lo biológico que enfatizan aspectos como las hormonas sexuales o las estructuras cerebrales, es menester dar cuenta de aspectos de socialización, entendiendo además la ordenación en categorías binarias durante el nacimiento y las recompensas (y sanciones) para los comportamientos que no se conformen a dichos géneros, de modo que personas de género femenino y masculino internalizan dichos comportamientos y producto de la socialización, prefieren (en su mayoría) comportamientos acorde a ello.

Desde la sociología, se distinguen explicaciones sobre estas desigualdades de género desde un nivel micro (es decir, las características de las personas) y un nivel macro (es decir, en términos de organizaciones y sociedades nacionales) y que además en su interacción se producen formas de desigualdad de género (Thébaud & Charles, 2018).

Sobre las explicaciones a nivel micro, destacan las relativas a las aptitudes, preferencias, prioridades, aspiraciones, expectativas y productividad laboral de las personas y las diferencias de género en dichas características (Cheryan et al., 2017; Stoet & Geary, 2018; Thébaud & Charles, 2018). Así, similar a lo observado por Stewart (2021), el género femenino observa una mayor orientación a las relaciones interpersonales y el cuidado a la par que el género masculino da cuenta de mayor capacidad para el pensamiento analítico. Sin embargo, las explicaciones expuestas no son en su generalidad vista como adecuadas, toda vez que, primero, dichos roles varían notoriamente a lo largo del tiempo y el espacio (Thébaud & Charles, 2018) e incluso dentro de las áreas STEM (Cheryan et al., 2017) y, segundo, porque las diferencias medibles son demasiado reducidas como para dar cuenta de los patrones de extrema segregación observados en muchas ocupaciones y lugares de trabajo (Thébaud & Charles, 2018).

Además de ello, la literatura da cuenta de que las predisposiciones no sólo no son idénticas entre individuos, sino que, además, no son capturadas correctamente en el sistema de género binario, de tal modo que incluso aspectos como la división de labores familiares no tienen gran capacidad explicativa en las diferencias en trayectorias de mujeres en áreas STEM respecto a áreas no STEM. Así, al parecer, dichas características individuales tienen una capacidad explicativa limitada en la comprensión de trayectorias diferenciadas por género. Sin embargo, Thebaud & Charles (2018) da cuenta de que sistemas de creencias esencialistas (como la percepción de que diferentes géneros tienen predisposiciones naturales a ciertas áreas), si es un factor causal relevante en la construcción de un sesgo cognitivo que explica las aspiraciones ocupacionales diferenciadas.

Luego, sobre las explicaciones a nivel macro, Thebaud & Charles (2018) da cuenta de cómo la reestructuración posindustrial de la economía gestó efectos desiguales en la estratificación por género, aumentando la tasa de ocupación femenina a la vez que contribuyó a la consolidación de núcleos de labores tipificadamente femeninas, generando una incorporación laboral como mujeres que no considera a la mayoría de las ocupaciones STEM como propia del género femenino. Dichos patrones, además, moldean las políticas y tradiciones nacionales en aspectos como las condiciones laborales, fuero familiar, cuidado de menores, hacia la reproducción de modelos normativos de trabajo y familia (Thébaud & Charles, 2018).

A continuación, un segundo conjunto de características de nivel macro, dan cuenta de cómo la creciente burocratización de las relaciones laborales ha obscurecido las definiciones sobre el género ideal de contratación, a la vez que dicho ideal permanece sesgado hacia lo masculino y heteronormativo (Cheryan et al., 2017).

Los puntos antes mencionados, sin embargo, no dan cuenta por completo de la segregación al interior de las áreas STEM si no se considera el género como un “marco organizativo de las relaciones sociales e identidades individuales que se inscribe en las instituciones económicas y sociales” (Thébaud & Charles, 2018, p. 8) y que de ello se desprende la segregación por género de las áreas ocupacionales, por lo que, a pesar de factores como las condiciones operativas (como la escasez de trabajadores o el desarrollo tecnológico), una vez que, así como los individuos, ciertos roles y áreas son marcados como propios de lo masculino o

lo femenino, ello es naturalizado por las personas y absorbido en la socialización a temprana edad (Cheryan et al., 2017). Un estudio experimental comentado por Thebaud & Charles (2018) da cuenta de cómo la exposición (falsa) de información sobre superioridad masculina en el desempeño en una tarea específica resultó en que personas de género femenino informasen una menor autoevaluación de desempeño en dicha tarea.

Entonces, en la interacción de los tres aspectos ya mencionados se dibuja una relación de triada en que, en paralelo y dialógicamente, las percepciones individuales sobre ciertas labores y áreas del conocimiento sesgan las pretensiones de un género u otro hacia ellas, lo que construye y alimenta estereotipos de género sobre dichos espacios que, en sentido opuesto, moldean las percepciones de las personas respecto a sus aptitudes propias y ajenas, contribuyendo adicionalmente a la discriminación en el reclutamiento (Charlesworth & Banaji, 2019; Thébaud & Charles, 2018) y en los espacios de trabajo respecto a participantes no percibidos como propios de un rol producto de su género (Charlesworth & Banaji, 2019; Moss-Racusin et al., 2018; Thébaud & Charles, 2018).

## 5 Marco metodológico

La presente investigación es de carácter cuantitativo, con una aproximación estadística y longitudinal basada en datos de fuentes secundarias, principalmente administrativas y recopiladas desde diferentes entidades relativas a la administración educativa nacional. La unidad de información es el estudiante y, para ello, cada conjunto de datos está individualizado mediante el RUN del estudiante enmascarado para asegurar un grado de anonimato en la provisión de información.

Los conjuntos de datos empleados se describen en la Tabla 1, comentando también los rangos anuales que describen y su origen.

**Tabla 1. Conjuntos de datos disponibles**

<b>Conjunto</b>	<b>Descripción</b>	<b>Fuente</b>	<b>Períodos</b>
Matrícula escolar	Matrícula anual en educación escolar primaria y secundaria. Datos sensibles provistos mediante solicitud de información al Ministerio de Educación.	Centro de Estudios MINEDUC	2004 - 2022
Rendimiento escolar	Desempeño y estado final anual del estudiante.	Centro de Estudios MINEDUC - Datos abiertos	2004 - 2021
Directorio de establecimientos	Registro de establecimientos escolares disponibles.		
Egreso escolar	Registro anual de egresos del sistema escolar mediante modalidad NNJ, PJA o alternativas remediales.		
Defunciones de estudiantes	Registro histórico de defunciones asociadas a estudiantes registrados en el sistema escolar.	Centro de Estudios MINEDUC / Registro Civil	2004 - 2022
Cuestionario de padres	Resultados de encuesta de caracterización para padres y apoderados realizada en paralelo a las evaluaciones SIMCE.	Agencia de Calidad de la Educación	2004 - 2018
Desempeño SIMCE	Desempeño de estudiantes en evaluaciones SIMCE a nivel individual.		
Inscripción PSU/PAES	Inscripción y resultados en evaluaciones PSU/PAES. Para el caso de PSU, datos sensibles provistos a nivel individual mediante solicitud de información a DEMRE.	DEMRE / Centro de Estudios MINEDUC - Datos abiertos	2004 - 2022
Matrícula SUA	Matrícula en educación terciaria realizada mediante Sistema Único de Admisión. Para el caso de PSU, datos sensibles provistos a nivel individual mediante solicitud de información a DEMRE.		

Conjunto	Descripción	Fuente	Períodos
Datos socioeconómicos SUA	Resultados de la encuesta de caracterización socioeconómica realizada durante la inscripción al Sistema Único de Admisión. Para el caso de PSU, datos sensibles provistos a nivel individual mediante solicitud de información a DEMRE.		
Postulaciones SUA	Postulación a programas de estudios mediante el Sistema Único de Admisión. Para el caso de PSU, datos sensibles provistos a nivel individual mediante solicitud de información a DEMRE.		
Matrícula terciaria	Matrícula anual en educación terciaria a nivel individual.	Sistema de Información de Educación Superior	2004 - 2022
Titulación terciaria	Títulos obtenidos en la educación terciaria informados anualmente.		2004 - 2021
Postulación a beneficios estudiantiles	Postulación a beneficios estatales. Incluye becas de arancel, gratuidad y fondo solidario de crédito universitario.		
Asignación de beneficios estudiantiles	Asignación de beneficios estatales producto de postulaciones.		
Asignación de Créditos	Registro histórico de créditos con aval del Estado / créditos de garantía estatal a nivel individual.	Comisión Ingresos	2004 - 2021

Fuente: Elaboración propia

## 5.1 Reconstrucción de trayectorias

La reconstrucción de trayectorias educativas es un procedimiento de análisis longitudinal que consiste en la elaboración de tablas temporalmente ordenadas que permiten trazar el paso por el sistema educativo de un estudiante para cada uno de los períodos desde un período inicial (sea este su ingreso al sistema o el primer registro disponible). Este procedimiento se sustenta e inspira en lo propuesto por Rodríguez y Villanueva(2022), esto es que mediante datos de carácter administrativo y longitudinal de nivel individual es posible identificar tendencias y patrones conducentes a eventos e hitos de interés y que, para ello, se producen dos productos principales, esto son las trayectorias ya previamente descritas y observaciones o resúmenes, siendo estos reducciones puntuales de la información acumulada, para permitir la descripción de trayectorias en un único caso. Para la realización de lo antes expuesto, se elabora un procedimiento de tres etapas.



### **5.1.1 Tablas de trayectorias escolares**

La primera etapa consiste en la elaboración del eje articulador sobre la matrícula individual de cada estudiante, para ello se identifica a cada uno de los estudiantes cuya trayectoria será reconstruida para luego emplear únicamente los datos elementales para la elaboración de una tabla de trayectorias inicial que describe únicamente los aspectos esenciales sobre su vínculo con el sistema educativo.

En el caso de la educación escolar, lo antes descrito corresponde a concatenar la matrícula disponible para cada año con el rendimiento relativo a este, sin embargo, es menester mencionar que el par matrícula-rendimiento no satisface un criterio de simetría ni de unicidad, es decir, ocurre que un estudiante puede presentar, para un mismo período, más de una matrícula y/o más de un reporte de rendimiento a la vez que una matrícula puede no disponer de un reporte de rendimiento o viceversa.

A modo de salvaguardar la completitud de la información, se realiza un mecanismo que, por tanto, considera a la vez la matrícula y el rendimiento anual para la reconstrucción de las trayectorias. Para ello se identifican todos los pares disponibles y, toda vez que uno de los dos datos disponibles no disponga de pareja con un dato en el sentido opuesto este se reconstruye, declarando un dato adicional que dicha reconstrucción fue llevada a cabo para posibles evaluaciones posteriores.

Lo anterior, por tanto, implica que si para una matrícula no existe un rendimiento equivalente (esto es para el mismo establecimiento educativo y estudiante), dicho rendimiento es generado sintéticamente imputando un promedio general y asistencia nula, y asumiendo a la vez que dicha inconsistencia es producto de un retiro intra-año desde el establecimiento.

En dirección opuesta, cuando se presenta un reporte de rendimiento sin una matrícula para ser emparejado, dicha matrícula es producida sintéticamente imputando valores nulos en los datos personales no disponibles en dicho reporte, como por ejemplo la comuna de residencia del estudiante, su situación de integración, entre otros.

Finalmente, para completar dicha etapa, se anexan los datos a una tabla ordenada que dispone de información nula para cada año desde que el estudiante está matriculado por primera vez y hasta que egresa por primera vez<sup>14</sup>, fallece o hasta el último año informado.

El producto del procedimiento antes descrito se denomina tablas de trayectorias y con ello se cierra la primera de las tres etapas de la reconstrucción para luego realizar la segunda etapa que consiste en la construcción de tablas de estado.

### **5.1.2 Tablas de estado escolar**

A diferencia de las tablas de trayectorias, que disponen de información para cada registro disponible del estudiante, aunque exista más de uno por período, las tablas de estado describen la situación del estudiante para un período de manera unívoca, esto quiere decir que existen tantos registros del estudiante como períodos de los que se disponga información (es decir, desde el comienzo de su trayectoria reconstruida hasta el final) con solo un registro de estado por período.

Esta etapa obliga a la desambiguación de las trayectorias, es decir, toda vez que para un mismo período se deba identificar una variable de interés (como, por ejemplo, el nivel), es necesario declarar un mecanismo de desambiguación del nivel que discrimine cual de todos los niveles a los que se asocia al estudiante en un mismo año es el de mayor relevancia.

En el caso de la educación escolar, dicha desambiguación consiste en identificar una matrícula principal por año en función de los siguientes criterios:

1. Matrícula en educación para niños, niñas y jóvenes.
2. Presentar aprobación o, en su defecto, repitencia.
3. Promedio general más alto respecto a otras matrículas del mismo estudiante en dicho año.
4. Presenta una matrícula en el establecimiento el año previo.
5. Presenta una matrícula en el establecimiento el año posterior.

---

<sup>14</sup> Considerando que existen estudiantes que presentan más de un egreso.

Dichos criterios son aplicados de manera ordenada hasta que no existan empates y, si ello no ocurre, se realiza un desempate aleatorio.

### **5.1.3 Tabla de observaciones escolares**

Finalmente, la tercera etapa consiste en la alimentación de las trayectorias con el resto de datos disponibles (nivel socioeconómico, desempeño a nivel escolar, entre otros), para luego realizar resúmenes (también llamados observaciones) sobre los datos que permiten acotar toda la información disponible sobre una trayectoria en variables de interés que describen de manera puntual aspectos de relevancia en la trayectoria del estudiante como, por ejemplo, recuento de sucesos como repitencias, períodos de deserción o indicadores binarios como si existió o no cambios de escuela o carrera terciaria.

En oposición, si bien es posible realizar un procedimiento análogo para los datos de educación terciaria a partir de los datos de matrícula y titulación de los estudiantes, ello implica confrontar el problema del alto volumen de titulaciones sin matrícula.

Por lo anterior, se propone una forma simplificada alternativa que consiste en considerar únicamente la matrícula en la reconstrucción de trayectorias terciarias y luego, incluir su intersección con la titulación en relación con su carrera año y, adicionalmente, reconstruyendo una matrícula si y sólo si el estudiante presenta una matrícula asociada a dicha carrera en un año igual o anterior.

### **5.1.4 Codificación individual de programas de estudios en nivel superior**

Para realizar lo antes comentado, es primordial atender la identificación de la situación del estudiante en una carrera, pues si bien en los datos administrativos se asocia un programa de estudios con un código único, esto requiere la corrección de los eventos en que se presentan estudiantes matriculados sin asociar a un código único o que un programa de estudios (identificado por nombre e institución) dispone de más de un código único.

Para lo anterior, la aproximación consiste en:

1. Identificar todas las carreras debidamente identificadas por su código único<sup>15</sup>.
2. Identificar, para cada año, a los estudiantes matriculados en una carrera sin código único.
3. Cuantificar el flujo de estudiantes que desde una carrera sin código único en un año  $t$ , están matriculados en una carrera con código único en  $t+1$  y dicha carrera está en la misma institución.

Entonces, se asume que la carrera de destino y la de origen en realidad son una única carrera si se cumplen todos los siguientes requisitos:

- La carrera sin codificar de origen y la carrera de destino corresponden a un mismo nivel global (pregrado, posgrado o postítulo).
- El flujo es de al menos 10 estudiantes.
- El flujo corresponde a un 30% de los matriculados en el origen o en el destino.
- Ambas carreras corresponden a una misma área genérica o disponen de una distancia de *Levenshtein*<sup>16</sup> mayor a 0,5.

Luego, si una carrera sigue sin disponer de un código de identificación único, entonces se le imputa un código único adicional y se suma que ese par nombre de carrera e institución es único y no coincide con ningún otro.

### **5.1.5 Tabla de trayectorias terciarias**

Además de lo antes descrito, a diferencia de las trayectorias escolares y en consideración de lo heterogéneo de la educación terciaria, se reconstruyen sub-trayectorias para cada programa en que el estudiante está asociado cada año, considerando la agregación de dichas sub-

---

<sup>15</sup> Toda vez que existe una carrera asociada a más de un código, se selecciona arbitrariamente uno de estos para preservar el carácter único de la asociación entre nombre y código.

<sup>16</sup> La distancia de Levenshtein, distancia de edición o distancia entre palabras es el número mínimo de operaciones requeridas para transformar una cadena de caracteres en otra- Esto para atender casos de menciones u otros programas conducidos desde un programa inicial y cuya área genérica difiere.

trayectorias como la trayectoria terciaria del estudiante y, por tanto, excluyendo la necesidad de generar tablas de estado, pero multiplicando la cantidad de observaciones por estudiante, siendo ahora una observación por estudiante y programa de estudios individualizado. Esto implica, además, que cada observación puede considerar, además, información sobre estudios posteriores y ulteriores.

Adicionalmente es necesario indicar algunas consideraciones para la estimación de la graduación y la graduación oportuna. En primer lugar, el completar una carrera no es un evento exento de consideraciones metodológicas, pues, por ejemplo, las licenciaturas conducentes a título pueden presentar más de una titulación y, adicionalmente, se presentan trayectorias con más de una titulación o que presentan matrícula luego de obtener títulos.

Para ello, la manera de abordar la completación de estudios asociados a un programa consiste en, primero, identificar, en función del título, si este consiste en una licenciatura o un título. Luego, si consiste en una licenciatura, se considera únicamente la primera de ellas y se descartan las posteriores y, si consiste en un título, se acota la trayectoria hasta el año de la titulación, excluyendo toda matrícula posterior a ella.

Una vez realizado dicho procedimiento, se realiza una estimación de la duración del programa de estudios, asumiendo el supuesto simplificador de que todo programa de estudios tiene una duración ideal asociada al nivel de la carrera, como se en la Tabla 2.

**Tabla 2. Duración teórica por tipo de carrera**

<b>Nivel de carrera</b>	<b>Duración teórica (en años)</b>
Técnico de nivel superior	3 (±1)
Profesional con licenciatura	5 (±1)
Profesional sin licenciatura	4 (±1)
Ciclo inicial o plan común <sup>17</sup>	2 (±1)
Licenciatura no conducente a título	5 (±1)

<sup>17</sup> Indiferente de considerarlo, los títulos de bachiller u otros ciclos iniciales no informan titulación a través del SIES.

<b>Nivel de carrera</b>	<b>Duración teórica (en años)</b>
Magíster	2 (+1)
Doctorado	5 ( $\pm$ 1)
Postítulo	1 (+1)
Especialidad del área de salud	3 ( $\pm$ 1)
Diplomado (superior a un semestre)	1 (+1)

Fuente: Elaboración propia

Y, que el rango de opciones de la duración se concreta considerando la moda de la duración real de estudios para cada programa, considerando el punto medio de la duración teórica toda vez que dicho valor exceda el rango establecido.

Con dicha consideración, se puede identificar el egreso a tiempo a partir de los años que el estudiante ha estado dentro del programa de estudios, sin embargo, es necesario considerar que: a) La estimación de duración de los programas es simplificada, b) Los estudiantes pueden realizar estrategias para adelantar años de estudios como, por ejemplo, la convalidación de asignaturas y c) La realización de ciclos iniciales es propia de programas de estudio heterogéneos que dificultan la correcta estimación de los programas que son consecución a partir de ellos. Por ello, se asume que el egreso a tiempo es en realidad una estimación cuya certeza es difícil de evaluar solo a partir de información administrativa.

Adicionalmente, considerando el tercer punto del apartado anterior, se estableció que, si se cumple, a la vez, que un estudiante cursó anteriormente un ciclo inicial en la misma institución y que dicho ciclo inicial pertenece a la misma área del conocimiento que una carrera posterior, entonces los años dentro del programa anterior se suman a los del actual y que es ese resultado el que se emplea en la estimación de duración ideal y, por tanto, de graduación a tiempo de los estudiantes.

A partir de la duración ideal declarada y la estimación tiempos, se determinó además identificar los retiros, entendido como estar matriculado en un programa de estudios identificado por un código único en un período  $t$  y no estarlo en el  $t + 1$ , aunque se pueda estar vinculado a un programa de estudios diferente.

Ello además permite generar una clasificación de retiros en inicial, toda vez que ocurre el primer año, temprano cuando ocurre en el segundo o tercer año en carreras de duración de cuatro o más años y solo en el segundo en caso opuesto, y tardío cuando ocurre en cualquier período posterior.

## **5.2 Alcance muestral**

En relación con la intersección en los períodos temporales para los que se dispone información de cada conjunto de datos y para asegurar un rango de trazabilidad de los estudiantes de una extensión razonable, se propone una investigación centrada en cohortes de estudiantes definida en función de su egreso escolar.

Para ello, considerando una trazabilidad de al menos siete años y, considerando que el último período con información completa es 2021, se define el análisis de las siete cohortes de egreso entre los años 2008 y 2014. Esto corresponde a 1.623.837 estudiantes, de quienes 823.710 (50,73%) son de género femenino y 800.127 (49,27%) son hombres. Esta elección de cohortes tiene un fin práctico, pues permite disponer de, al menos, 8 años para trazar trayectorias educativas e identificar la incidencia de fenómenos relativos a ella.

Al respecto, es impropio referirse al alcance muestral de esta investigación, toda vez que se dispone de información completa y cuasi-censal de la población de estudio delimitada, ello producto principalmente del empleo de datos administrativos que disponen información total sobre el sistema educativo de nivel superior y sus estudiantes en relación con su matrícula, titulación u otros eventos relacionados.

## **5.3 Procedimientos de análisis exploratorio guiado**

Para analizar la existencia de brechas de género en las trayectorias educativas en nivel superior, se realiza una aproximación desde el análisis exploratorio de datos (EDA, por sus siglas en inglés) sustentado en aprendizaje automático. El análisis exploratorio de datos es una etapa en el descubrimiento de conocimientos en la ciencia de datos que consiste en la realización secuencial de análisis, pero que a su vez requiere tanto de habilidades analíticas como también conocimiento del dominio (Milo & Somech, 2020).

Para simplificar los aspectos más complejos del EDA, como los altos costos en tiempo de realización y las numerosas líneas de trabajo y aristas que emergen al aproximarse a un nuevo conjunto de datos, se emplea una aproximación guiada por aprendizaje automático que consiste en utilizar diferentes modelos de clasificación o regresión para divisar potenciales aristas de interés, reduciendo la cantidad de operaciones de exploración y, a la vez, divisando esquemas iniciales de los resultados de manera simplificada para orientar de manera eficiente los aspectos de interés a evaluar en el interior de los datos disponibles (Milo & Somech, 2020). Para complementar el planteamiento anterior, se propone emplear los valores SHAP (Lundberg et al., 2018) para identificar la relevancia global de cada descriptor en los fenómenos de interés y luego, dado el foco sobre género que se expone en la presente investigación, utilizar el índice de Shapley-Taylor (Dhamdhere et al., 2020), una aproximación de segundo orden para la estimación de los valores SHAP, para identificar cuán relevante es dicho descriptor por sí mismo y en interacción con otros descriptores.

Los algoritmos de aprendizaje automático a emplear corresponden en su totalidad a la familia de los algoritmos de árboles de decisión, específicamente con potenciación de gradiente y con *binning*, tales como *XGBoost*, *LightGBM*, *CatBoost* (Bentéjac et al., 2021) y una implementación específica en la biblioteca de scikit-learn llamado *Histogram Based Gradient Boosting Classifier* inspirada en *LightGBM*<sup>18</sup>.

Los algoritmos de árbol de decisión segmentan los ejercicios de clasificación en espacios decisionales en forma de árbol de ramas binarias<sup>19</sup>, cuya selección de descriptor para particionar y el valor específico de segmentación dependen de la ganancia de información para el algoritmo entendida esta como la información mutua que provee una variable observada respecto a otra de interés (Bentéjac et al., 2021). El uso de *boosting* es una estrategia para reducir el *overfitting* mediante la concatenación ponderada de sub-árboles (también llamados *weak learners*) que se agregan para obtener una predicción final, omitiendo la captación completa del fenómeno en cada uno de ellos y mejorando así su desempeño en el eje *bias-variance* (Bentéjac et al., 2021) y,

---

<sup>18</sup> Más detalles en: <https://scikit-learn.org/stable/modules/ensemble.html#histogram-based-gradient-boosting>

<sup>19</sup> Excepto *XGBoost* cuyas ramas son ternarias para lidiar con valores nulos.



por su parte, el *binning* consiste en la partición de los descriptores continuos en segmentos discretos para optimizar los tiempos de selección del punto de partición decisional.

La ventaja de estos algoritmos es, primero, la ausencia de requerimientos sobre el comportamiento de los predictores respecto al predictor, no asumiendo formas específicas de la relación y, por tanto, permitiendo la captación de descriptores de cualquier nivel de medición sin requerir formas de *encoding* para su inclusión correcta en ellos<sup>20</sup>. Adicionalmente, dada la naturaleza intrínseca del algoritmo, excluye de las particiones aquellas variables que, estando fuertemente correlacionadas, contribuyan la misma información al modelo, preservando únicamente aquellas cuyo aporte en el ejercicio de clasificación sea de mayor prioridad en cuanto a su ganancia de información.

El desempeño de estos modelos, en tanto modelos de clasificación, se realiza a partir de sus predicciones por clase, para ello, considerando los resultados que son verdaderos positivos (VP), verdaderos negativos (VN), falsos positivos (FP) y falsos negativos (FN) se calculan las siguientes métricas:

- *Precision* (clase positiva):  $\frac{VP}{VP+FP}$
- *Precision* (clase negativa):  $\frac{VN}{VN+FN}$
- *Recall* (clase positiva):  $\frac{VP}{VP+FN}$
- *Recall* (clase negativa):  $\frac{VN}{VN+FP}$
- Puntaje de media geométrica:  $\sqrt{\text{Recall}(\text{clase positiva}) * \text{Recall}(\text{clase negativa})}$

De estas, el puntaje de media geométrica es el seleccionado para evaluar y seleccionar al algoritmo de mejor desempeño, toda vez que su formula equilibra el error por clase del modelo y, por tanto, su resultado es indiferente del desbalance de clases el fenómeno a describir (Rodríguez & Villanueva, 2022).

---

<sup>20</sup> Aunque ello puede simplificar su interpretación.

Por su parte, los valores SHAP son una herramienta analítica de estimación de los valores Shapley, siendo estos una propuesta basada en teoría de juegos para evaluar la contribución marginal de diferentes actores sobre la ganancia generada por una actividad (Lundberg et al., 2018). La estimación de los valores SHAP se realiza haciendo perturbaciones locales de los valores de cada variable para estimar la contribución de cada predictor para cada predicción individual (Ribeiro et al., 2016). La propuesta de SHAP para estimar los valores Shapley satisfacen criterios de *local accuracy*, *meaningness* y *consistency* (Lundberg et al., 2019). Las principales implicaciones de estas tres características son, primero, el carácter aditivo de los valores SHAP, esto significa que, de modo similares modelos de fácil explicabilidad (como una regresión de mínimos cuadrados), el resultado predicho se computa directamente en función de los valores obtenidos (en el caso de SHAP, como la suma del valor base más el valor SHAP de cada predictor) (Lundberg et al., 2018).

Además, una segunda aproximación a los valores SHAP, el índice de Shapley-Taylor, es una implementación de segundo orden en la estimación de los valores Shapley, que permite desglosar los resultados en forma de una matriz que expresa en su diagonal la contribución unitaria de cada descriptor y en sus no diagonales la contribución de cada descriptor en interacción con otros (Dhamdhere et al., 2020), ello inhibe la necesidad de emplear aproximaciones alternativas a la medición de la interacción como la dependencia parcial o el análisis de varianza de dos o más vías.

Se propone una distinción entre descriptores significativos y relevantes basada en dos criterios. El primer criterio indica que una variable es significativa si y sólo si su contribución al modelo es significativamente mejor que un descriptor aleatorio. Para ello se emplea el algoritmo de Boruta (Kursa et al., 2010), que compara en múltiples iteraciones al descriptor respecto a una versión aleatoriamente permutada de sí mismo, lo que permite evaluar si su contribución es significativa o si, en cambio, su presencia es descartable. Luego, un segundo criterio, para contribuir a la parsimonia de los análisis a realizar es que un descriptor es relevante si y sólo si su contribución al modelo es igual o mayor al 5%. Esto en términos globales, considerando que la unidad de los valores SHAP en modelos de clasificación binaria es el aporte a la probabilidad de pertenencia a la clase positiva, significa que un descriptor es relevante sólo si su importancia

SHAP<sup>21</sup> es mayor a 0,05 y, en términos relativos, en los análisis de la interacción entre descriptores, se considera a una interacción relevante sólo si su contribución en el aporte total de dicho descriptor es igual o mayor al 5% de este.

### **5.3.1 Hitos identificados**

En línea con lo realizado por Blanco (2018), se propone un análisis orientado por hitos, en que cada hito se define como un evento realizado por un estudiante y que es relevante para su trayectoria educativa.

Se propone, de manera similar a Blanco (2018), consideración de los siguientes hitos:

- a) Egreso de educación escolar. Esta elección de cohortes tiene un fin práctico, pues permite disponer de, al menos, 8 años para trazar trayectorias educativas e identificar la incidencia de fenómenos relativos a ella.
- b) Postulación mediante Sistema Único de Admisión. (en los primeros tres años luego del egreso), que es posible segmentar en dos apartados. Primero, si las estudiantes postulan en igual cantidad y, segundo, si dicha postulación está orientada o no hacia áreas de menor feminización, específicamente aquellas que corresponden a la clasificación GEEMP, donde antecedentes como Bordón et al. (2020) informan que es esperable que existan notorias brechas en la intención de participar de las mujeres en estas áreas curriculares.
- c) Matrícula terciaria (primer registro), que es similar al ejercicio anterior, considerando la incidencia de la matrícula los tres años siguientes al egreso escolar y el área hacia la que están orientadas, donde también Bordón et al. (2020) sugiere que ella sea desigual respecto al género de las matrículas.

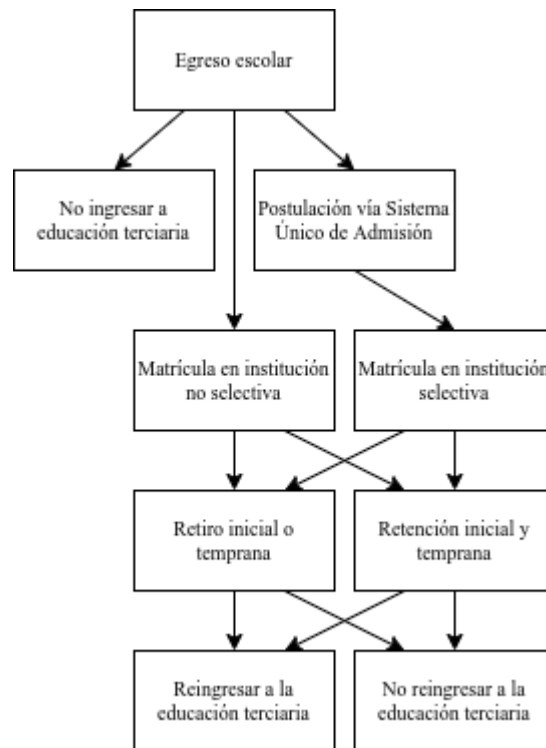
---

<sup>21</sup> La importancia SHAP es la media del valor absoluto de los valores SHAP calculados por cada descriptor.

- d) Retención inicial y temprana, la que se espera sea desigual en áreas GEEMP, considerando que la revisión de literatura sugiere la existencia de factores que inciden en una menor retención femenina en dicha área (Thébaud & Charles, 2018), considerando solo retiro temprano y tardío.
- e) Transición entre programas, también llamada reingreso, para la que se consideran tres años posteriores al retiro y se espera que, si se sigue lo informado en la literatura (Richardson et al., 2020; Stewart-Williams & Halsey, 2021; Thébaud & Charles, 2018), la presencia de discriminación hacia las mujeres en áreas masculinizadas u otros factores que incidan en su retiro devienen en un tránsito hacia áreas no masculinizadas.

Como se observa, si bien cada uno de los hitos es requisito necesario para el siguiente (salvo b), la realización del siguiente hito es una salida opcional frente a otras alternativas dispuestas y válidas dentro del sistema educativo. Una visión general de ello se propone en el siguiente diagrama.

**Figura 13. Diagrama decisional de trayectorias por hito**



Fuente: Elaboración propia

Como se visualiza en la Figura 13, la opción de desistir en la participar de la educación terciaria, dado su carácter no obligatorio, está siempre presente y, por tanto, debe ser considerada en los diferentes análisis a realizar.

Es menester considerar que, además de las hipótesis por hito antes expuesta, se busca identificar potenciales interacciones entre la incidencia de cada fenómeno en que se exprese desigualdad de género y cómo ello potencialmente se relaciona con otros descriptores, a fin de identificar potenciales subgrupos dentro de cada hito que sean afectados en mayor o menor magnitud por este, tal y cómo comentan autores como Blanco et al. (2018) o Manquepillan & Sepúlveda (2017) al hacer notar que si bien el género es un factor relevante en la desigualdad educativa, este incide en paralelo y en conjunto a otros notorios aspectos como son, principalmente, el nivel socioeconómico u otros elementos como el tipo de institución o programa educativo que se cursa.

En lo aplicado, cada uno de los hitos será analizado desde la óptica de su ocurrencia, las condiciones y manera en que ello ocurre, la incidencia del género en ello y cómo esta característica interactúa con las otras características del hito. Para el caso puntual de la titulación y/o completación de estudios, se determinó que dada la información incompleta y estimativa de duración de los programas, y la heterogeneidad propia de las trayectorias terciarias, sumado a que ciertas categorías no informan sus egresos y que, además, las cohortes seleccionadas disponen de solo ocho años de registros en el peor de los casos, es preferible enfocarse en los apartados iniciales de la trayectoria y omitir, frente a potenciales sub-reportes, la consideración por aquellos tardíos.

Retomando el punto previo, se desglosará cada hito en dos etapas intercomunicadas. Primero, una fase de análisis descriptivo para dar cuenta de aspectos diferenciadores en la manera en que cada hito ocurre y, segundo, una etapa de análisis basada en un modelo de aprendizaje automático que se ajuste sobre los descriptores de ocurrencia de cada evento, para así analizar el resultado de sus predicciones descomponiendo la contribución marginal de cada variable en la predicción e identificar aspectos de relevancia para explorar con detalle en la fase descriptiva. La contribución marginal de cada variable se calculará mediante valores SHAP.

Es menester hacer notar dos aspectos sustantivos en la elección metodológica expuesta. En primer lugar, se prefiere emplear los valores SHAP por sobre otras aproximaciones como el uso de regresiones lineales en conjunto a otras estimaciones de términos de interacción (como la dependencia parcial o los análisis de varianza), dado el carácter agnóstico de los valores SHAP y su aplicabilidad a cualquier modelo predictivo, indiferente de los parámetros y requerimientos específicos de este, permitiendo una mayor flexibilidad en la consideración de predictores con diferentes niveles de medición; y, segundo, a pesar del empleo de modelos de aprendizaje automático, la presente investigación carece de ánimos predictivos y el uso de estos modelos satisfacen nada más que la necesidad de un dispositivo que capte las diferencias en múltiples descriptores sobre la ocurrencia de un fenómeno, de modo que aspectos como la división de muestra para entrenamiento y validación u otros aspectos similares se apegan a aproximaciones tradicionales para su realización, toda vez que la predicción final y sus resultados son indiferentes para las intenciones analíticas aquí expuestas.

#### 5.4 Declaración de descriptores

A continuación, en la Tabla 3, se declaran todos los descriptores a emplear a lo largo del análisis, ellos fueron clasificados en dimensiones a partir de los comentarios de Thébaud & Charles (2018) y Blanco et al. (2018), considerando cómo los antecedentes sugieren que dichos descriptores y dimensiones son relevantes a la hora de delimitar diferencias en trayectorias postsecundarias y cómo estas interactúan con el género del estudiante.

Además, de acuerdo con lo expuesto en la revisión de literatura, se identifica como relevante la distinción de las áreas GEEMP y LPS dentro del área STEM, es por ello por lo que en los análisis por área de estudios se reclasificaron los programas de acuerdo a su clasificación CINE según se informa en la Tabla 15.

*Tabla 3. Descriptores por dimensión*

<b>Dimensión</b>	<b>Descriptor</b>	<b>Observaciones</b>
Sociodemográfica	Género	Reclasificada en forma binaria donde 1 es género femenino
	Edad de egreso	
	Comuna de residencia	Código Único Territorial (CUT) de la comuna de residencia en el egreso

<b>Dimensión</b>	<b>Descriptor</b>	<b>Observaciones</b>
Desempeño	Promedio escolar	Promedio escolar de la enseñanza media. El promedio general de cada año fue estandarizado respecto al contexto del estudiante por nivel y establecimiento.
	Puntaje PSU/PAES Lenguaje	Puntaje obtenido en cada prueba específica. Estandarizados para simplificar su comparabilidad entre años y entre evaluaciones.
	Puntaje PSU/PAES Matemáticas	
Escolar	Tipo de enseñanza	Científico-Humanista, Técnico-Profesional o Artística. Reclasificada como una serie de descriptores binarios.
	Modalidad	Egreso desde educación para niños, niñas y jóvenes (NNJ) o en modalidad para jóvenes y adultos (PJA). Reclasificada como una serie de descriptores binarios.
	Dependencia	Dependencia de egreso: pública, particular subvencionado o particular pagado sin aportes estatales. Reclasificada como una serie de descriptores binarios. Los establecimientos de administración delegada se consideran subvencionados. Dadas las cohortes seleccionadas, no se registran egresos desde servicios locales de educación.
	Ruralidad del establecimiento	Informado en forma binaria por el Ministerio de Educación.
Institucional	Categoría de institución	Categorizada en institutos técnicos, centros de formación técnica, universidades privadas no participantes del SUA y universidades adscritas al CRUCH o privadas participantes del SUA según se destaca en la revisión de antecedentes. Reclasificada como una serie de descriptores binarios.
	Categoría del programa de estudios	Categorizado en programas profesionales sin licenciatura, licenciaturas no conducentes a título, profesionales con licenciatura, técnico de nivel superior y plan común, ciclo inicial o bachillerato. Reclasificada como una serie de descriptores binarios.
	Clasificación STEM	Clasificación STEM basada en una reclasificación de la subárea CINE 2013 según se detalla en la Tabla 15. Reclasificada como una serie de descriptores binarios.
Económica	Nivel socioeconómico	Nivel socioeconómico estimado basado en el índice de vulnerabilidad multidimensional del establecimiento y el ingreso declarado del estudiante según se detalla en el anexo B.
	Postulación a Beneficios Estatales (BE)	
	Beneficio Estatal recibido	Clasificada en becas de arancel, gratuidad y créditos. Reclasificada como una serie de descriptores binarios.
Específicas (trayectoria)	Año de egreso	
	Años desde el egreso	

Fuente: Elaboración propia

## 6 Resultados

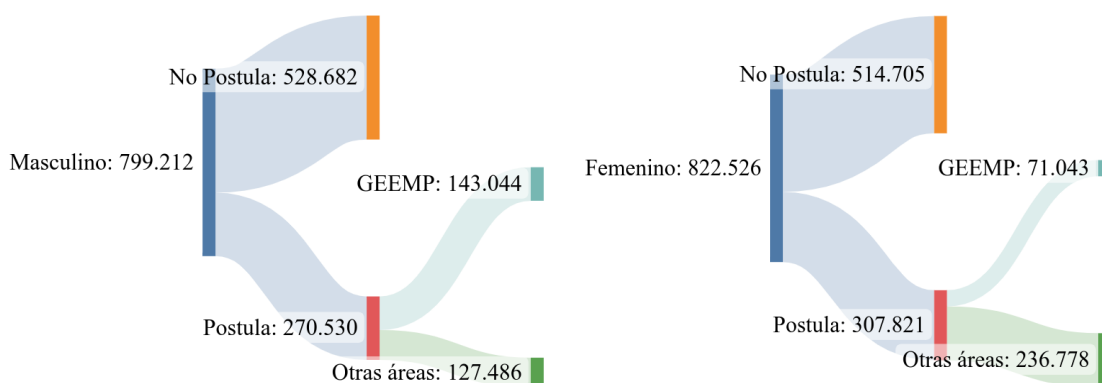
A continuación, se exponen los resultados del análisis exploratorio guiado antes propuesto. En primer lugar, se expone un resumen del panorama general de las trayectorias educativas con relación al género y, luego, se expone el detalle de los resultados de los análisis en relación con cada hito indicado.

### 6.1 Panorama general de trayectorias en educación superior por género

Se trazan las trayectorias de 1.612.738 estudiantes que egresan de la educación escolar entre 2008 y 2014. Por género, estos son 822.526 (50,72%) estudiantes de género femenino y 799.212 (49,28%) masculino. De ellos, 1.020.180 (62,91%) egresan desde la enseñanza media científico-humanista, 601.464 (37,09%) desde la enseñanza técnico-profesional y solo 94 desde la enseñanza de tipo artística (0,005%). Además 85,1% egresa desde educación regular para niños y jóvenes, esto son 1.380.053 de los estudiantes, mientras que los otros 241.685 egresan desde educación para jóvenes y adultos (PJA).

Así, como se observa en la Figura 14, si bien el egreso y la postulación son aspectos más o menos similares en cuanto al género, la postulación femenina hacia el área GEEMP mediante el SUA es notoriamente inferior, con una tasa de postulación respecto al total de postulantes de 23,08% respecto al género masculino cuya tasa es de 52,88%.

Figura 14. Diagrama sankey de postulación por género

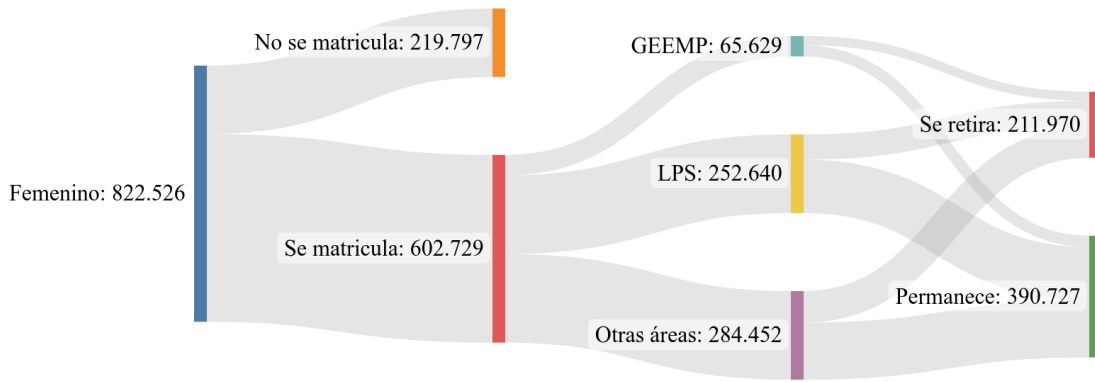


Fuente: Elaboración propia



Luego, al visualizar las trayectorias que expone la Figura 15, podemos notar que la matrícula femenina, respecto al total de egresados es de 73,28%, y que de ellas un 10,89% se matrícula en GEEMP, mientras un 41,92% en LPS y un 47,19% en otras áreas. La tasa de retiro inicial o temprano para cada una de estas áreas es de 45,04%, 32,09% y 35,63% respectivamente.

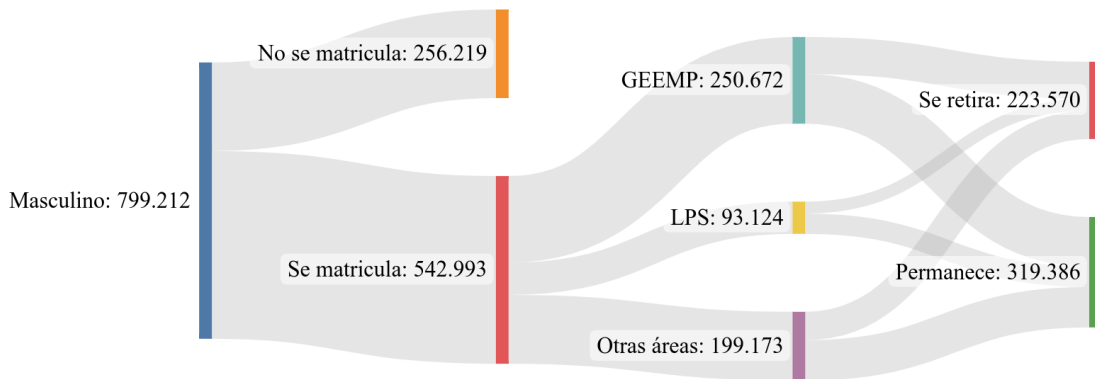
**Figura 15. Diagrama sankey de tránsitos del género femenino desde la matrícula hasta el retiro**



Fuente: Elaboración propia

En cuanto al género masculino, dichas cifras son de una tasa de matrícula de 67,94%, que se distribuye en un 46,16% hacia el área GEEMP, un 17,15% hacia el área LPS y un 36,68% hacia otras áreas, con una tasa de retiro inicial o temprano de 42,67%, 37,14% y 41,18% respectivamente, tal como se observa en la Figura 16.

**Figura 16. Diagrama sankey de tránsitos del género masculino desde la matrícula hasta el retiro**

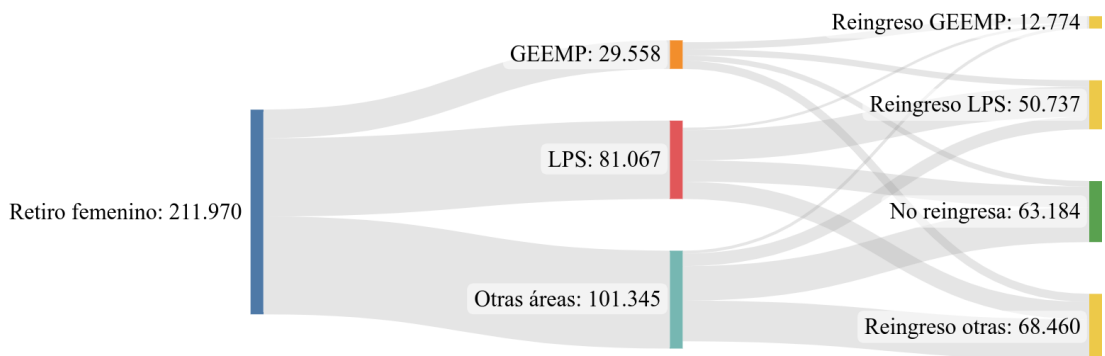


Fuente: Elaboración propia

Así podemos visualizar que mientras la matrícula masculina en GEEMP es proporcionalmente mayor a la femenina, también lo es la retención, esto implica no solo un filtro durante el ingreso que, por cualquier causa, excluye a las mujeres de estas áreas, sino también un filtro durante las carreras que las excluye en mayor medida respecto a otras áreas y también respecto al género masculino.

Además de ello, de estos retiros, podemos observar en la Figura 17 que el reingreso femenino, que corresponde al 65,27% de los retiros, se distribuye tal que un 13,94% desde GEEMP, 38,24% desde LPS y 47,81% proveniente de otras áreas, con una permanencia intra-área (es decir, estudiantes que se retiran y reingresan en una misma área, considerando a los no reingresos) de 24,23% para el área GEEMP, 38,67% para LPS y 41,56% para otras áreas.

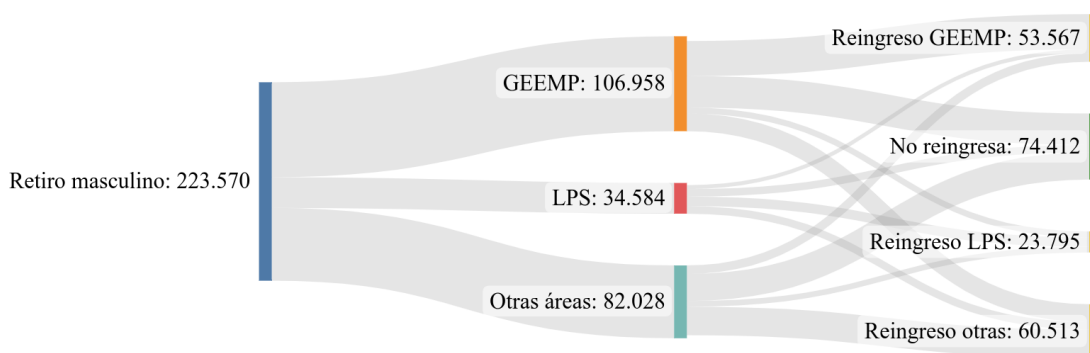
**Figura 17. Diagrama sankey de tránsito desde retiro a reingreso para el género femenino**



Fuente: Elaboración propia

Para el caso masculino, se visualiza en la Figura 18 el reingreso general es de un 64,90%, que se distribuye en 47,84% desde GEEMP, 15,47% desde LPS y 36,69% desde otras áreas, con una permanencia intra-área de 36,99% para el área GEEMP, 30,93% para LPS y 39,02% hacia otras áreas.

Figura 18. Diagrama sankey de tránsitos desde retiro a reingreso para el género masculino



Fuente: Elaboración propia

Lo anterior da cuenta de un notorio sesgo en las trayectorias de educación superior específicamente en la elección de carrera exhibe una recurrente subrepresentación femenina en el área GEEMP junto con una ligera sobrerrepresentación en el área LPS, sin embargo, para dar cuenta de cómo dicho fenómeno se nutre de su interacción con otros aspectos, se procede al análisis exploratorio de datos guiado por aprendizaje automático, como ya fue expuesto en la metodología presentada.

Los resultados específicos y detallados de cada uno de los modelos a continuación expuestos se detallan en el Anexo C.

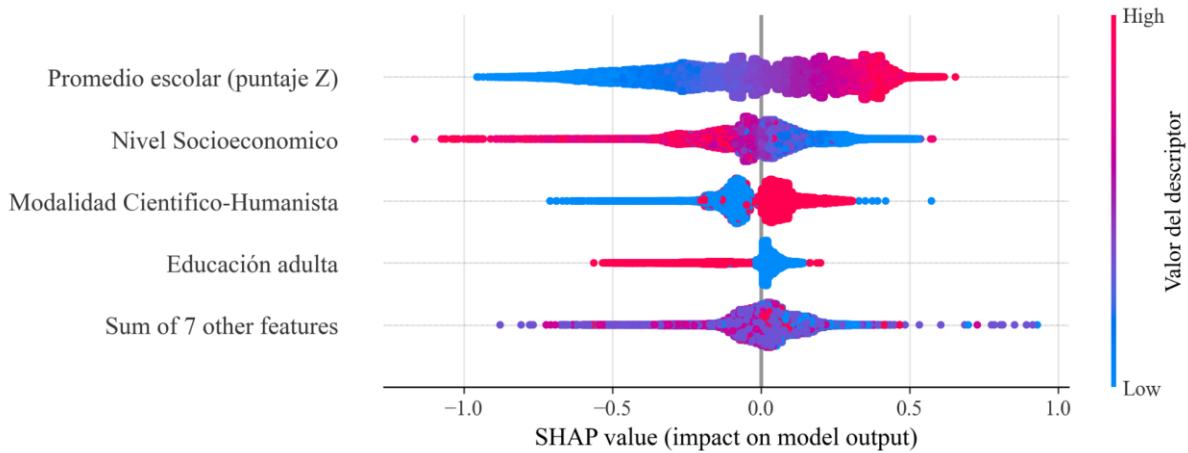
## 6.2 Egreso escolar

Para el egreso escolar se consideraron los 12 descriptores ya antes mencionados para identificar tendencias que permitan comprender diferencias sustanciales en estudiantes femeninos y masculinos. El algoritmo de *XGBoost* fue seleccionado con un puntaje de media geométrica de 0,578, que nos indica que, pese a existir diferencias, la capacidad clasificatoria del modelo es reducida, de lo que es posible inferir que ambos grupos son relativamente homogéneos.

Al respecto, como se observa en la Figura 19, los descriptores de mayor relevancia para distinguir el egreso de un estudiante masculino respecto a uno femenino son si provienen o no de

educación adulta, de modalidad científico-humanista, su nivel socioeconómico y su promedio escolar estandarizado.

**Figura 19. Diagrama de enjambre para los valores SHAP del primer modelo-hito<sup>22</sup>**



Fuente: Elaboración propia

Esto considerando únicamente los descriptores con una importancia SHAP igual o mayor a 0,05 como se describe en la Tabla 4.

**Tabla 4. Importancia SHAP de los descriptores del primer modelo-hito**

Descriptor	Importancia	Descriptor	Importancia
Promedio escolar (puntaje Z)	0,550	Modalidad Técnico-Profesional	0,022
Nivel socioeconómico	0,114	Dependencia Part. Subvencionada	0,019
Modalidad Científico-Humanista	0,087	Dependencia Part. Pagada	0,016
Educación adulta	0,052	Establecimiento rural	0,016
Edad de egreso	0,032	Dependencia Pública	0,015
Año	0,032		

Fuente: Elaboración propia

<sup>22</sup> Cada punto representa un conglomerado de uno o más casos, en el esquema azul-rosa se representa la media del descriptor para cada punto. El eje X expresa la contribución hacia la probabilidad de incidencia del caso positivo.

Adicionalmente, se evalúan los efectos de interacción para todos los descriptores relevantes, como se expone en la Figura 20.

**Figura 20. Mapa de calor de los valores Shapley-Taylor estimados para el primer modelo-hito**

Dependencia Part. Pagada	0.032	0.002	0.013	0.000	0.000	0.041	0.002	0.000	0.000	0.001	0.016
Promedio escolar (puntaje Z)	0.002	0.261	0.037	0.010	0.004	0.021	0.010	0.004	0.010	0.009	0.035
Ed. Adulta	0.013	0.037	0.041	0.005	0.001	0.021	0.017	0.001	0.006	0.014	0.030
Dependencia Part. Subvencionada	0.000	0.010	0.005	0.013	0.004	0.027	0.005	0.000	0.003	0.004	0.023
Establecimiento rural	0.000	0.004	0.001	0.004	0.015	0.004	0.002	0.000	0.002	0.001	0.011
Modalidad Científico-Humanista	0.041	0.021	0.021	0.027	0.004	0.070	0.011	0.005	0.000	0.009	0.075
Año	0.002	0.010	0.017	0.005	0.002	0.011	0.026	0.006	0.004	0.004	0.019
Dependencia Pública	0.000	0.004	0.001	0.000	0.000	0.005	0.006	0.014	0.001	0.002	0.018
Modalidad Técnico-Profesional	0.000	0.010	0.006	0.003	0.002	0.000	0.004	0.001	0.017	0.002	0.015
Edad de egreso	0.001	0.009	0.014	0.004	0.001	0.009	0.004	0.002	0.002	0.032	0.012
Nivel socioeconómico	0.016	0.035	0.030	0.023	0.011	0.075	0.019	0.018	0.015	0.012	0.146
	Dependencia Part. Pagada	Promedio escolar (puntaje Z)	Ed. Adulta	Dependencia Part. Subvencionada	Establecimiento rural	Modalidad Científico-Humanista	Año	Dependencia Pública	Modalidad Técnico-Profesional	Edad de egreso	Nivel socioeconómico

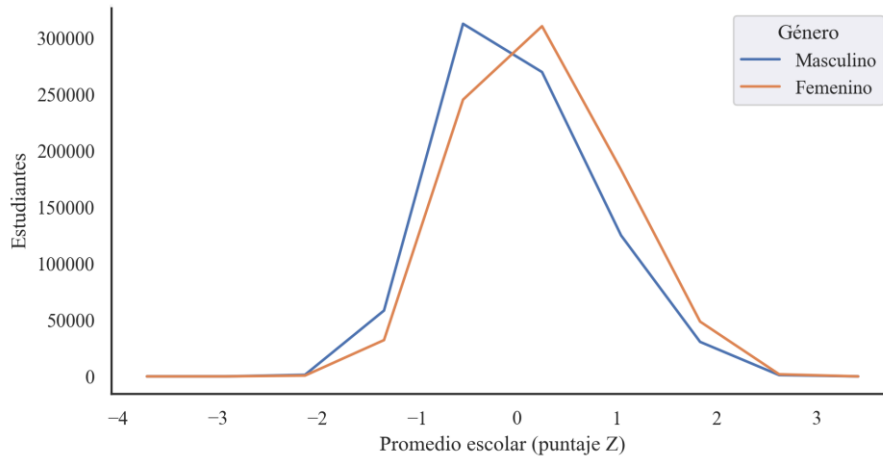
Fuente: Elaboración propia

Así, podemos observar cómo existe una interacción entre pertenecer a la modalidad científico-humanista en el egreso con el nivel socioeconómico, así como también existe interacción entre el nivel socioeconómico y el promedio escolar y, finalmente, entre pertenecer o no a la educación adulta con el promedio escolar y el nivel socioeconómico.

Así, como se observa en la Figura 21, el egreso femenino se caracteriza por disponer de un desempeño relativo ligeramente superior, aun dentro de la primera desviación estándar respecto a sus pares, pero con un notorio punto de corte en el primero, es decir, la mayoría de las

mujeres se encuentran por encima de la media de desempeño de sus pares, mientras que la mayoría de los hombres se encuentran bajo este umbral.

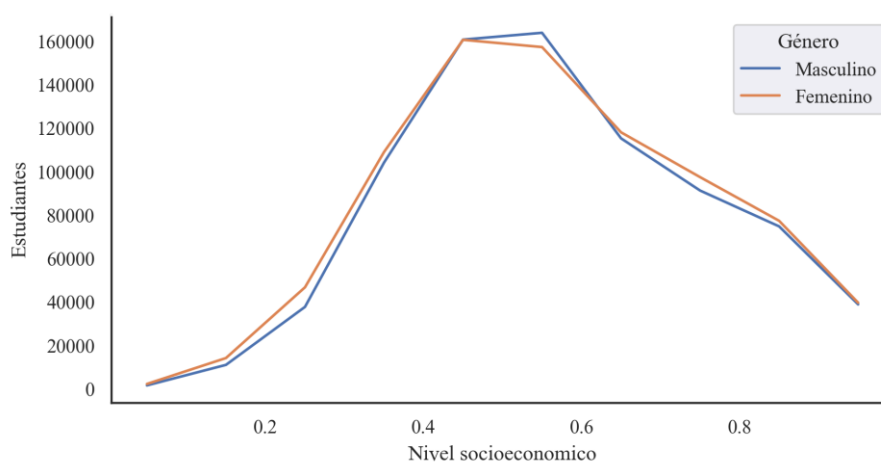
**Figura 21. Distribución por género del promedio escolar en la población estudiada**



Fuente: Elaboración propia

Luego, con respecto al nivel socioeconómico, podemos observar en la Figura 22 cómo su relevancia unitaria en la diferenciación no es evidente al visualizar las distribuciones para ambos géneros, viendo como ambas se concentran la media y mediana de la distribución con tendencia ligeramente asimétrica concentrada en el sector derecho (de mayor ingreso) de la distribución.

**Figura 22. Distribución por género del nivel socioeconómico en la población estudiada**



Fuente: Elaboración propia

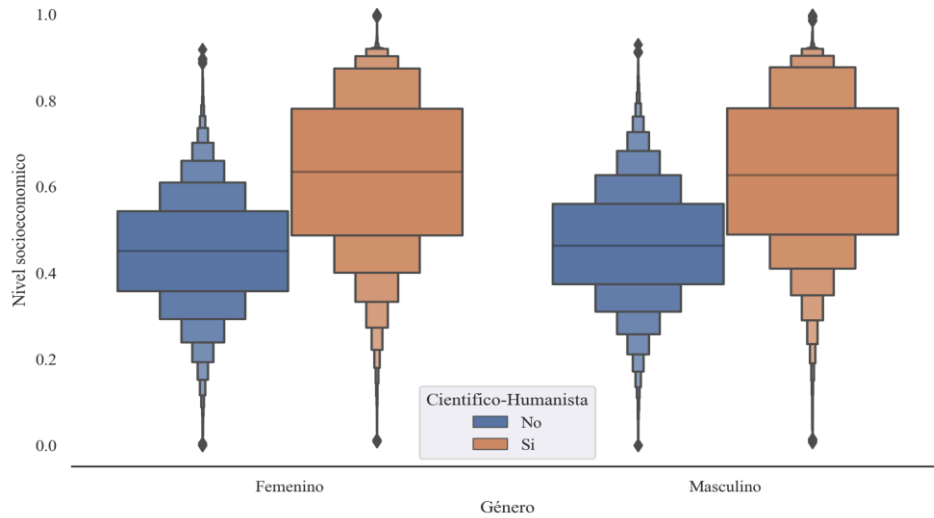
En sintonía con dicho hallazgo, la relevancia de egresar desde educación media científico-humanista también se observa relevante unitariamente, observando cómo a nivel poblacional la distribución por género es de 49,28% de estudiantes masculinos y 50,72% femeninos, distribución similar a la observada en los egresados desde la forma Científico-Humanista, con un 48,68% (+0,4% respecto a la población) de estudiantes masculinos y 51,32% femeninos. Esto concuerda con una importancia SHAP de solo 0,086 del que el 87,5% corresponde a su impacto unitario (sin interacciones). Asimismo, ocurre para el caso de estudiantes en establecimientos de dependencia particular pagada sin subvención estatal, en que la proporción es de un 51,62% (+2,34% respecto a la población) de estudiantes masculinos y 48,38% femeninos.

Lo anterior, para el caso de la educación adulta, produce una mayor diferenciabilidad, notando que, de los matriculados en educación adulta que registran egreso en las cohortes estudiadas, un 57,78% es de género masculino y el restante 42,24% es de género femenino, es decir, respecto a la población, existe una subrepresentación de 8,48% de estudiantes de género femenino.

Sin embargo, al visualizar en la Figura 23 la relación con los efectos de interacción antes mencionados, se observa que, primero, los estudiantes en modalidad Científico-Humanista exhiben un nivel socioeconómico más concentrado hacia los segmentos de mayores ingresos, a la

par que, de estos, se observa una ligera separación que indica que en dicha modalidad hay mayor cantidad de estudiantes femeninas con un mayor nivel socioeconómico.

Figura 23. Gráfico de *letter-value* del nivel socioeconómico por modalidad de egreso desglosada por género



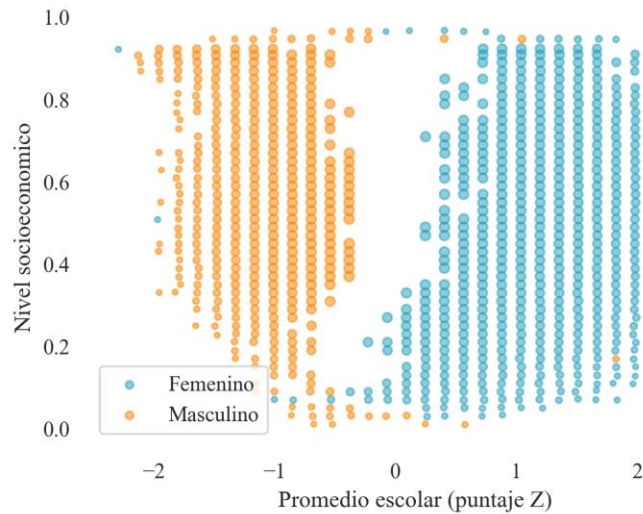
Fuente: Elaboración propia

Sin embargo, pese a ser relevante, dicha diferencia es de marginal al visualizar las medidas de tendencia central de ambas distribuciones, visualizando que la mayor diferencia se presenta en la mediana de los estudiantes que sí participan de educación científica humanista, con un 0,6269 de mediana en estudiantes de género femenino y 0,634 para estudiantes de género masculino, esto indica que, como es el caso de descriptores anteriores, la existencia de efectos de interacción significativos no implica que ello sea relevante.

Luego, respecto al nivel socioeconómico y su interacción con el promedio escolar estandarizado, podemos ver en la Figura 24 que, como ya observamos previamente, existe una distribución uniforme de nivel socioeconómico para ambos géneros y que el promedio escolar presenta diferencias de interés que favorecen al género femenino, sin embargo, además de lo expuesto, es posible observar que existe una ligera sobre participación femenina en los deciles inferiores del nivel socioeconómico.



Figura 24. Gráfico de dispersión segmentado de relación entre nivel socioeconómico y promedio escolar por género<sup>23</sup>



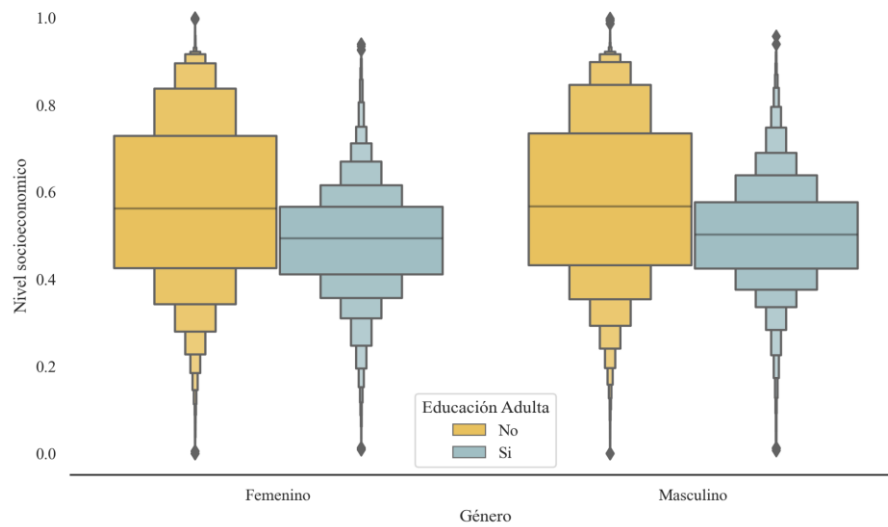
Fuente: Elaboración propia

Finalmente, respecto a los efectos de interacción observados para la educación adulta con respecto al nivel socioeconómico y el promedio escolar, el primero de estos efectos de interacción es marcadamente notorio, visualizando como, primero, existe una matrícula predominantemente masculina en la modalidad PJA, aspecto ya observado previamente, y segundo, es ligeramente observable que existe una diferenciación por género que indica que un estudiante en educación PJA y de bajo nivel socioeconómico es más probablemente masculino que femenino y, en sentido contrario, un estudiante en modalidad NNJ de alto nivel socioeconómico es probablemente femenino, aunque es destacable que dicho efecto de interacción es de sólo 0,03, por lo que la diferencia es prácticamente insignificante. Esto se observa en la .

---

<sup>23</sup> El gráfico está deliberadamente sesgado para identificar sobrerrepresentación femenina. Cada punto identifica un segmento de 50 particiones realizadas por cada eje, el tamaño del círculo es la magnitud de estudiantes en su interior y el color se asigna en función del género mayoritario. Solo se dibuja un círculo si la presencia de un género es al menos 10% mayor que el otro.

**Figura 25. Distribución de nivel socioeconómico por modalidad de egreso desglosado por género**

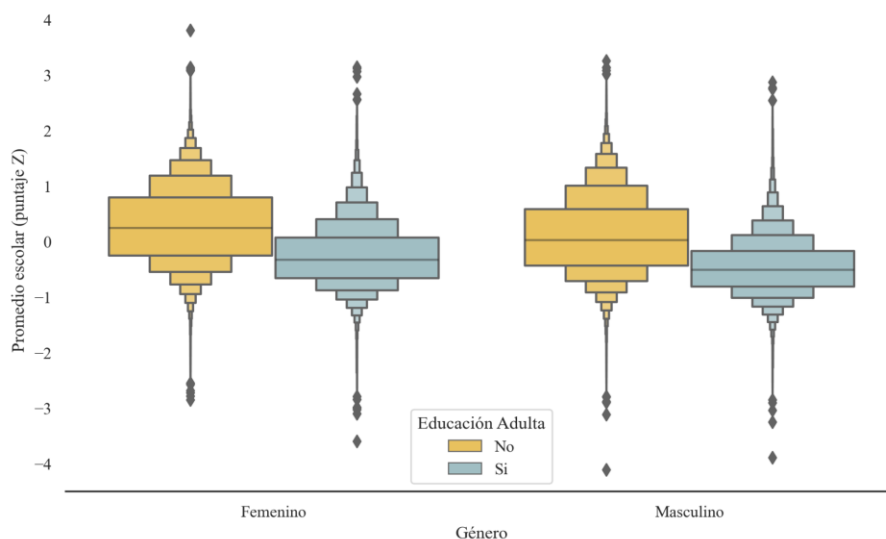


Fuente: Elaboración propia

Lo anterior se confirma al visualizar que la media y la mediana de las distribuciones para cada género presentan distancias de menos de 0,01.

Y, segundo, con un efecto de interacción de 0,037, es observable que, en su participación conjunta sobre la diferenciabilidad por género, los estudiantes en modalidad PJA de género masculino tienden a un promedio escolar más marcadamente por debajo de la media que las estudiantes de género femenino y, en sentido contrario, las estudiantes de género femenino en modalidad NNJ tienen a un promedio escolar por sobre la media.

**Figura 26. Distribución del promedio escolar por modalidad de egreso desglosado por género**



Fuente: Elaboración propia

En ello se expone una diferencia de mayor notoriedad, visualizando que los estudiantes de género masculino que no participan de educación adulta tienen un menor promedio escolar que las estudiantes, con una media de 0,106 y una mediana de 0,031, contra un 0,294 y 0,251 respectivamente, de manera similar que ocurre en la categoría opuesta, con una media de promedio escolar masculino de -0,458 y una mediana de -0,504, respecto a un -0,258 y -0,322 respectivamente. Es decir, hay diferencias entre categorías, donde los estudiantes que egresan de educación adulta presentan un menor promedio escolar, pero también existen diferencias intra-categoría, donde las estudiantes de género femenino, en ambas categorías, exhiben un mejor promedio escolar, con una diferencia más notoria en la categoría de estudiantes que egresan desde la modalidad NNJ.

### **6.3 Postulación al Sistema Único de Admisión**

Luego, respecto a las diferencias visibles en los estudiantes que egresan respecto a si postulan o no al Sistema Único de Admisión (luego de haber rendido la evaluación PSU o PAES), se consideran 12 descriptores, donde egresar de educación artística fue descartado al considerarse irrelevante en comparación a un descriptor aleatorio, se considera al algoritmo de *LightGBM* con un puntaje de media geométrica de 0,777, es decir, si bien dista de ser perfecto, el

modelo es capaz de discernir con relativa confianza que estudiantes postulan vía SUA en el momento del egreso.

En este caso, de los 12 predictores evaluados, 8 de ellos superan el umbral definido para considerarlos significativos, de los que el género de los estudiantes es el de menor importancia SHAP, como se expone en la tabla a continuación.

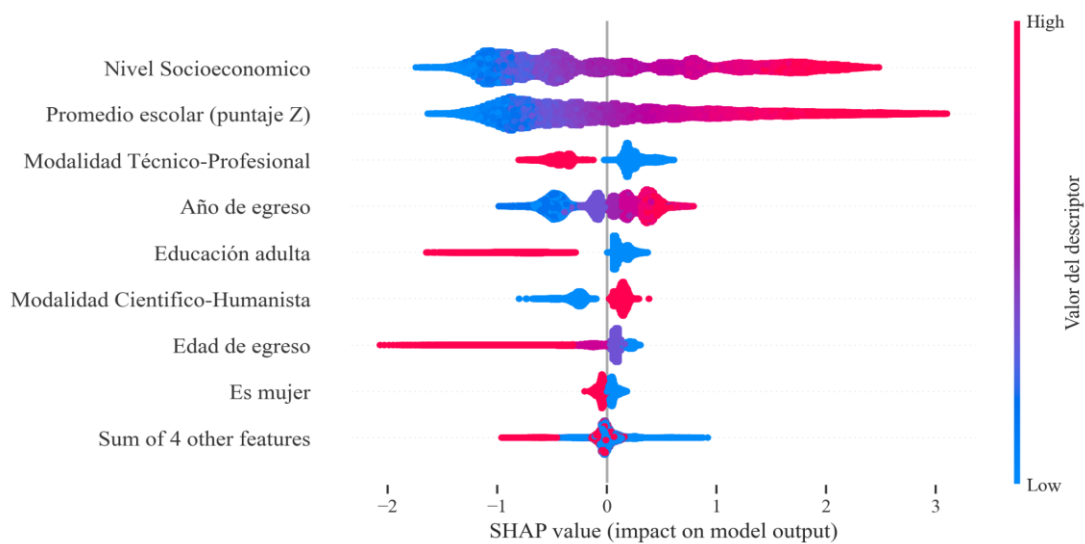
**Tabla 5. Importancia SHAP de los predictores del segundo modelo-hito**

<b>Descriptor</b>	<b>Importancia</b>	<b>Descriptor</b>	<b>Importancia</b>
Nivel Socioeconómico	0,879	Edad de egreso	0,157
Promedio escolar (puntaje Z)	0,755	Es mujer	0,056
Modalidad Técnico-Profesional	0,323	Dependencia Pública	0,035
Año de egreso	0,323	Dependencia Part. Pagada	0,035
Educación adulta	0,234	Establecimiento rural	0,027
Modalidad Científico-Humanista	0,189	Dependencia Part. Subvencionada	0,019

Fuente: Elaboración propia

La dirección de dicha relevancia se puede visualizar en términos generales en el gráfico de enjambre asociado, que nos indica que el nivel socioeconómico es el descriptor más relevante para dicha predicción junto al promedio escolar relativo al contexto del estudiante, ambos con un aporte positivo en su probabilidad estimada de postulación, en sentido opuesto a la modalidad técnico profesional que incide negativamente en dicha estimación y, de lo que nos es relevante, aparentemente el género femenino tiende a disponer de una menor probabilidad de postular, aunque ello con una relevancia casi descartable de solo 0,056, solo 6 milésimas por encima del umbral declarado.

**Figura 27. Diagrama de enjambre para los valores SHAP del segundo modelo-hito**

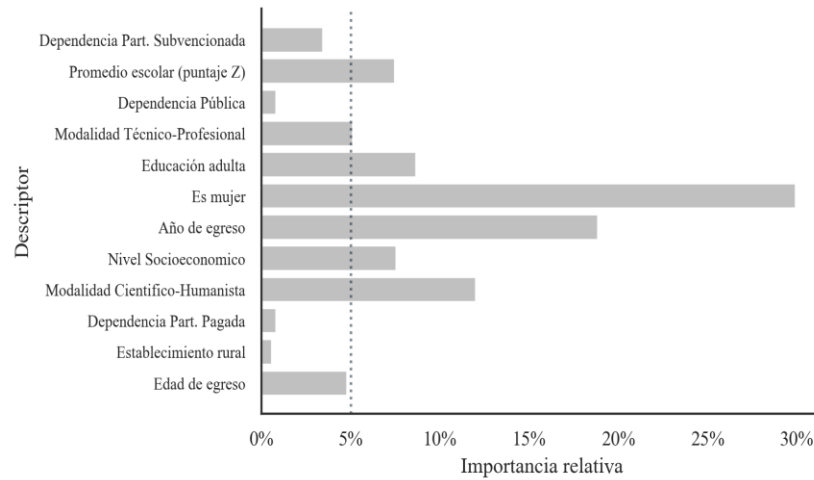


Fuente: Elaboración propia

De esta importancia, es destacable que únicamente un 30% de ella corresponde a la variable unitariamente, mientras el resto se distribuye, principalmente, en su interacción con el promedio escolar, egresar o no de modalidad PJA, el año de egreso, su nivel socioeconómico y si el tipo de educación media fue científico-humanista, como se visualiza a continuación.

En cuanto al género de manera unitaria, se observa que, si bien el 50,72% de la población estudiada es femenina, un 53,32% es postulante al SUA, lo que nos indica una ligera mayor disposición a postular para las estudiantes.

**Figura 28. Importancia relativa de los descriptores en su interacción con el género<sup>24</sup>**



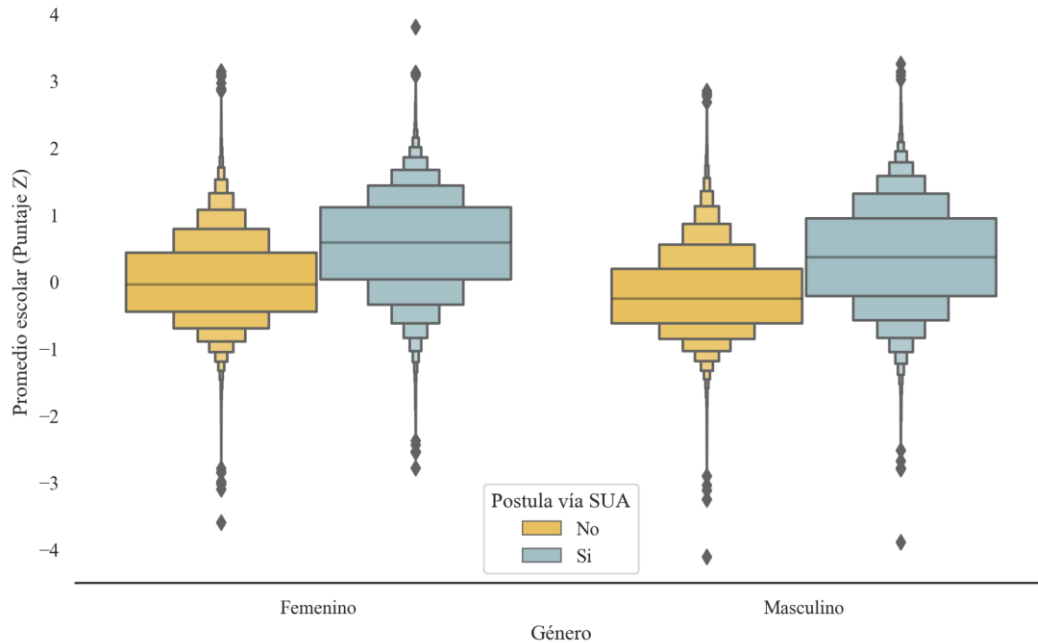
Fuente: Elaboración propia

Respecto al segundo de estos descriptores, se observa que, en general, la propensión a postular o no depende del promedio escolar estandarizado, sin embargo, se observa también que dicha diferencia se amplifica al contrastar entre ambos géneros, donde las estudiantes tienen una diferencia entre sus distribuciones de menor escala que los estudiantes, con una diferencia en su media de 0,547 entre quienes postulan y quienes no, respecto a la diferencia expuesta por los estudiantes de género masculino, con una distancia dos centésimas superior. Esto indica que, a pesar de que la población exhibe diferencias que el modelo considera significativas, su relevancia práctica permanece reducida.

---

<sup>24</sup> La importancia relativa es calculada como el porcentaje de contribución media de cada descriptor respecto a la contribución total de otro. En este caso el género.

**Figura 29. Distribución del promedio escolar por postulación vía SUA desglosado por género**

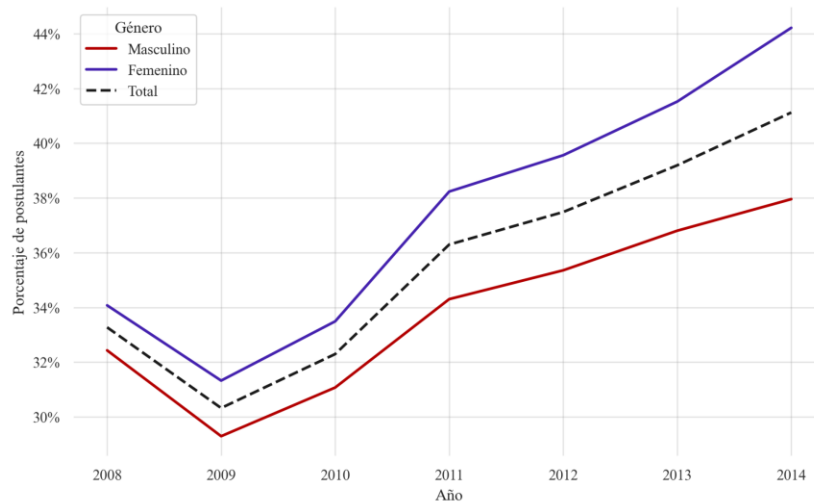


Fuente: Elaboración propia

Luego, respecto a la interacción entre el género y la educación adulta, podemos observar que el 40,7% de los egresados de dicha modalidad es de género femenino, respecto a un 59,3% de género masculino, mientras que quienes postulan mediante SUA son un 45,2% estudiantes de género femenino y 54,7% masculino. Esto indica que, existe una mayor probabilidad de que un egresado de modalidad PJA postule si es de género femenino que si es de género masculino.

Además, podemos observar que dicha disposición a postular ha aumentado con el tiempo, avanzando desde menos de un 34% en la cohorte de egreso 2008 hasta un 41% para la cohorte 2011.

**Figura 30. Evolución de tasa de postulación por cohorte de egreso**



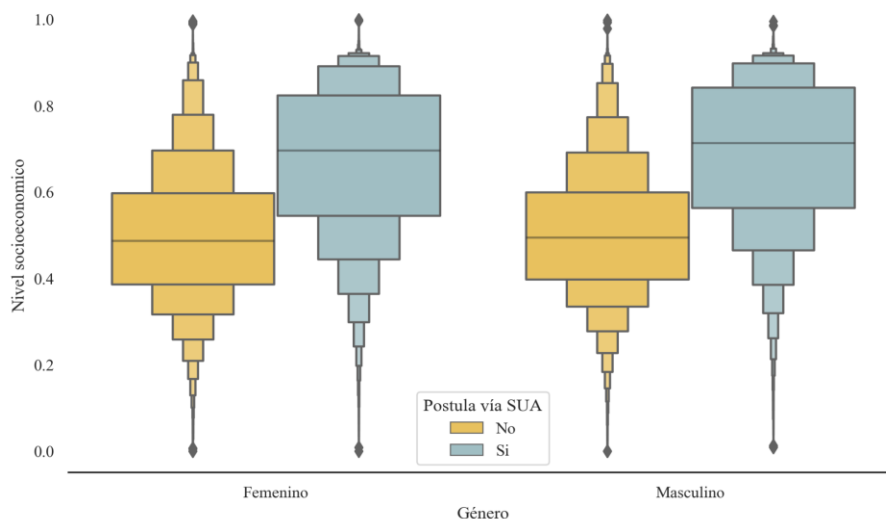
Fuente: Elaboración propia

Además, podemos visualizar como la distancia entre ambos géneros se ha acrecentado ligeramente con el tiempo, mostrando cómo dicha diferencia era de un 1% en la primera cohorte estudiada, lo que se ha duplicado hacia la última cohorte, lo que coincide con el hecho de que el 20% de la importancia del género como predictor de postulación corresponde a su interacción con la cohorte de ingreso.

Luego, respecto al nivel socioeconómico de los estudiantes, es visible como el de los postulantes se concentra notoriamente en el tramo superior de ingresos y, adicional a ello, se presenta una reducida diferencia entre ambos géneros, con una mayor distancia intra categoría en el género femenino, lo que indicaría una muy leve tendencia de estudiantes de género femenino a abstenerse de postular en relación a un menor nivel socioeconómico en comparación a los estudiantes de género masculino, aunque considerando a la vez la baja importancia del descriptor de género y una interacción sólo ligeramente superior al 7% en la contribución de dicha interacción, este efecto también resulta irrelevante.



Figura 31. Distribución del nivel socioeconómico por postulación vía SUA desglosado por género



Fuente: Elaboración propia

Finalmente, la contribución a la probabilidad estimada de egresar desde educación media científico-humanista respecto a postular vía SUA, considerando que solo un 42,2% de los egresados son de género femenino, nos muestra que estudiantes de dicho género están ligeramente más dispuestos a postular, con una presencia de 45,3% en los postulantes del SUA.

Como se observó previamente, la diferencia de mayor relevancia, en respaldo a lo comentado por la literatura, no es en cuanto al área STEM en su totalidad, sino específicamente en el segmento GEEMP, por ello se consideró también el sub-hito de postular o no a dicho área para los estudiantes que postulan mediante el Sistema Único de Admisión.

Para seleccionó al modelo entrenado con el algoritmo *CatBoost* que presenta un desempeño ligeramente mejor que el azar de 0,667 según su puntaje de media geométrica, con siete de sus doce predictores clasificados como relevantes, donde el género es el segundo descriptor que más contribuye luego del puntaje obtenido en la prueba específica de matemáticas en la PSU o PAES, con una importancia SHAP de 0,625 para la segunda y 0,62 para la primera, como se detalla en la tabla siguiente.

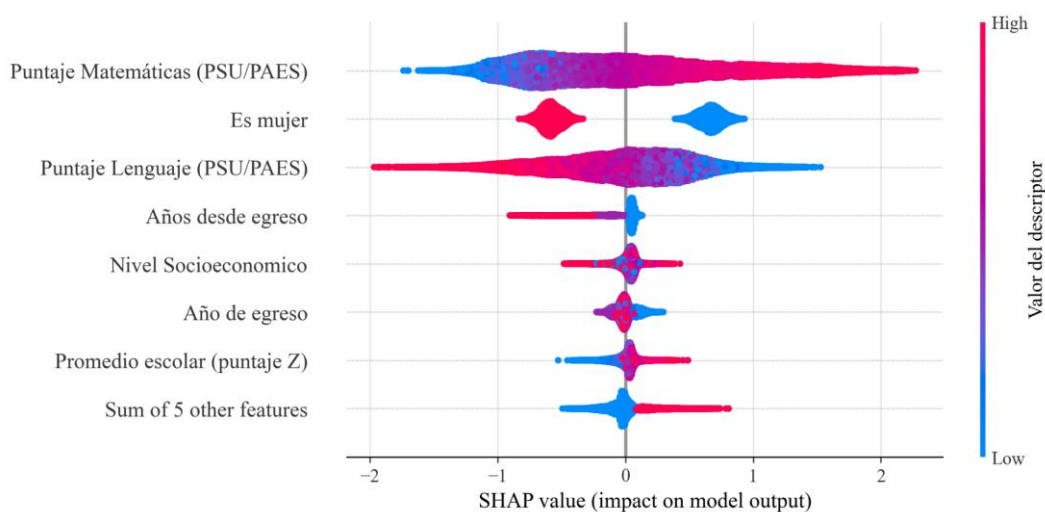
**Tabla 6. Importancia SHAP de los descriptores en el segundo hito - Submodelo de postulación hacia GEEMP**

Descriptor	Importancia	Descriptor	Importancia
Puntaje Matemáticas (PSU/PAES)	0,625	Promedio escolar (puntaje Z)	0,057
Es mujer	0,620	Modalidad Técnico-Profesional	0,047
Puntaje Lenguaje (PSU/PAES)	0,406	Dependencia Part. Pagada	0,043
Años desde egreso	0,084	Modalidad Científico-Humanista	0,028
Nivel socioeconómico	0,059	Dependencia Part. Subvencionada	0,012
Año de egreso	0,057	Edad de egreso	0,011

Fuente: Elaboración propia

En cuanto a la direccionalidad del aporte de los descriptores, podemos observar cómo el resultado en la prueba específica de matemáticas contribuye positivamente a la probabilidad de postular en el área STEM, ello en sentido opuesto a los resultados de la prueba específica de lenguaje y habilidades lectores, que inciden negativamente en dicha probabilidad. Respecto al género, existe una marcada distancia que muestra que las estudiantes de género femenino presentan una menor probabilidad de postular a dicha área.

**Figura 32. Gráfico de enjambre de valores SHAP para los descriptores del submodelo de postulación GEEMP**



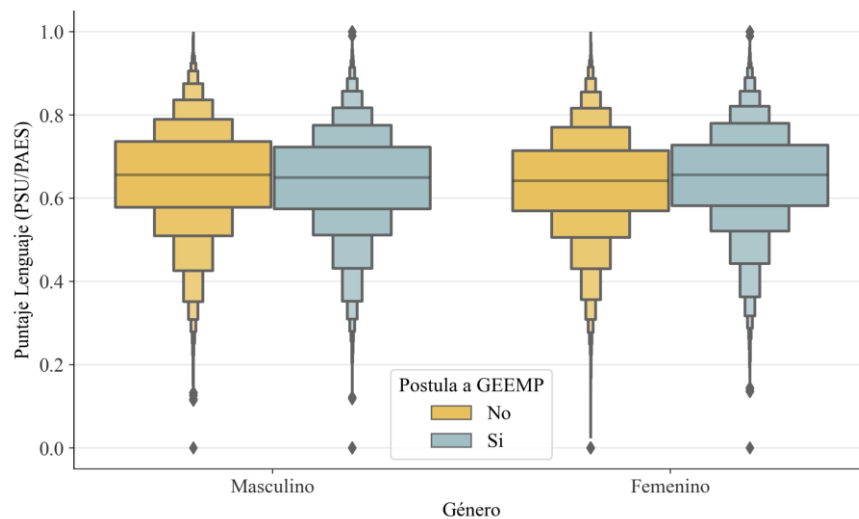
Fuente: Elaboración propia

Al respecto, es visible que, unitariamente, el género contribuye casi un 70% de su impacto, mientras el resto se distribuye principalmente en el resultado de ambas pruebas

específicas. Así, con una distribución en la población de postulantes que es 53,22% femenina, de este segmento solo un 33,18% postula al área GEEMP, mientras que el 66,82% restante es de género masculino.

En cuanto a las pruebas específicas, podemos visualizar que, primero, en lenguaje y en su interacción con género, para el caso de las estudiantes, su distribución de resultados en dicha evaluación es mayor en las postulantes respecto a las que no, situación opuesta al género masculino, en que el resultado de la evaluación mencionada es ligeramente inferior en los estudiantes que si postulan al área GEEMP. Sin embargo, dicha diferencia es casi irrelevante al observar la distribución en general.

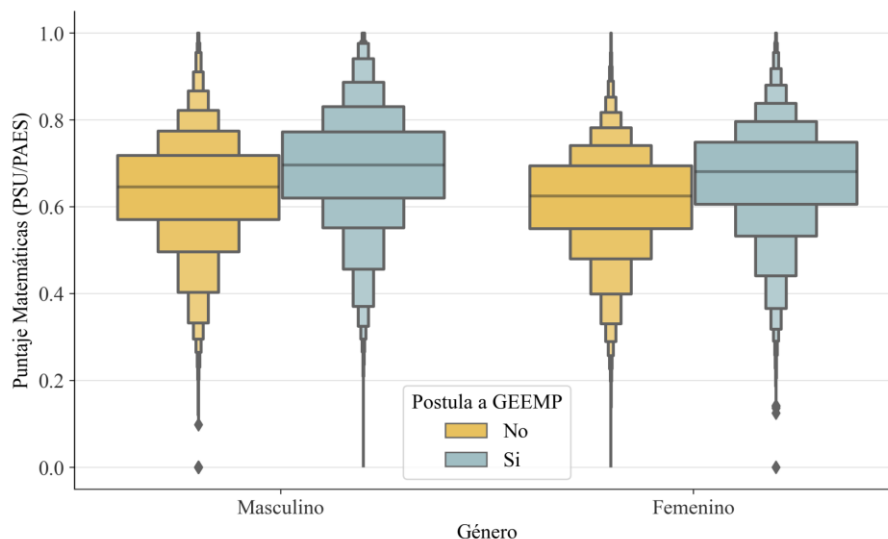
**Figura 33. Distribución del puntaje en la prueba PSU/PAES de lenguaje por postulación a GEEMP desglosado por género**



Fuente: Elaboración propia

En el caso de la evaluación específica de matemáticas, se observa una tendencia más homogénea entre grupos, donde el género femenino tiende a un resultado levemente inferior en sus puntajes en dicha prueba, lo que incide en que su disposición a postular tienda a tener una distribución también levemente más baja, aunque no se exhiben severas diferencias entre ambos géneros.

**Figura 34. Distribución del puntaje en la prueba PSU/PAES de matemáticas por postulación a GEEMP desglosado por género**



Fuente: Elaboración propia

## 6.4 Matrícula

Luego, en una ramificación que vuelve a los egresados, se evalúa qué aspectos con aquellos que permiten distinguir quienes se matriculan en educación superior y cuál es la incidencia del género en dicha revisión. Ello con un modelo entrenado con el algoritmo de *XGBoost* con un resultado similar al modelo antes descrito de 0,669 en su puntaje de media geométrica.

Nuevamente el género es un descriptor de poca relevancia, aunque supera el umbral declarado con una importancia SHAP de 0,068, pero muy distante de los descriptores más relevantes como son el nivel socioeconómico (con una importancia de 0,785) y el promedio escolar relativo (cuya importancia es de 0,421).

**Figura 35. Importancia SHAP de los descriptores del tercer modelo-hito**

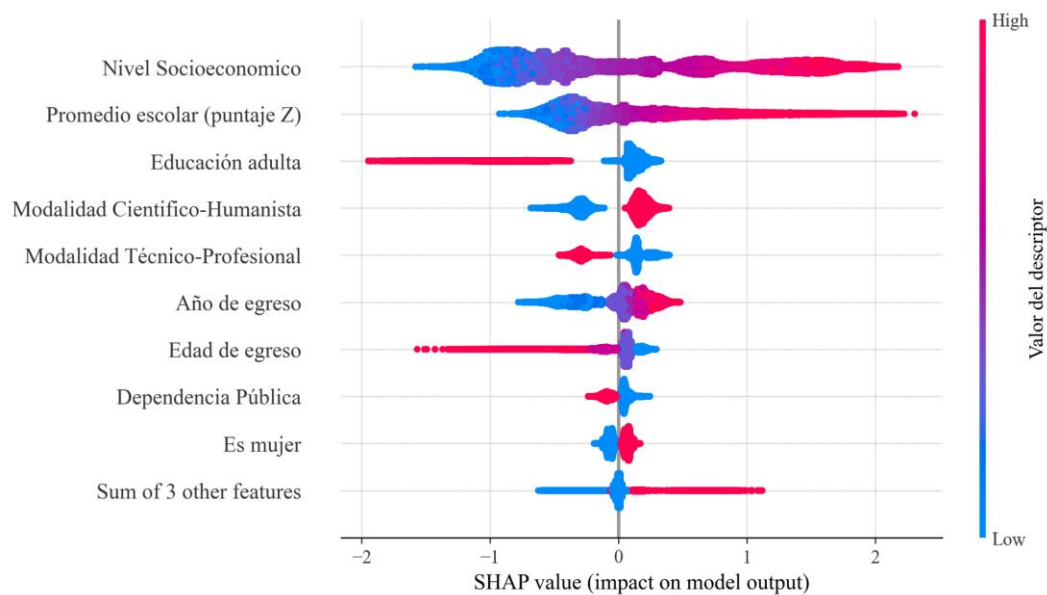
Descriptor	Importancia	Descriptor	Importancia
Nivel socioeconómico	0,785	Edad de egreso	0,120
Promedio escolar (puntaje Z)	0,421	Dependencia Pública	0,069
Educación adulta	0,245	Es mujer	0,069
Modalidad Científico-Humanista	0,230	Dependencia Part. Pagada	0,040

Descriptor	Importancia	Descriptor	Importancia
Modalidad Técnico-Profesional	0,218	Establecimiento rural	0,025
Año de egreso	0,206	Dependencia Part. Subvencionada	0,014

Fuente: Elaboración propia

Además, se observa que la incidencia del género es positiva para las estudiantes, esto quiere decir que ser mujer mejora levemente las probabilidades de que un estudiante se matricule en educación superior.

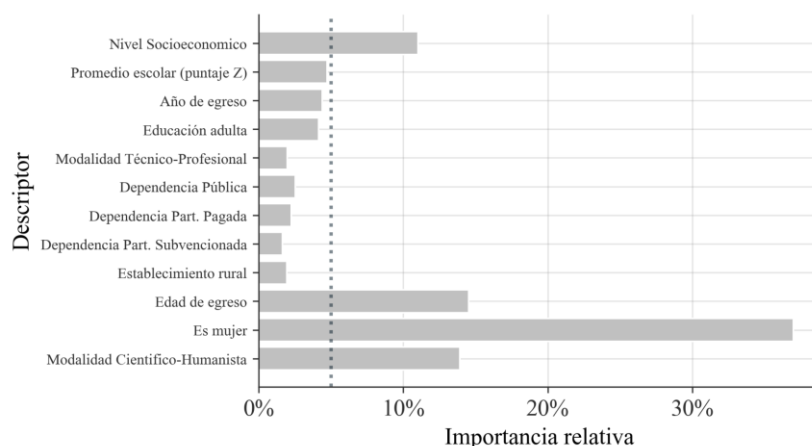
**Figura 36. Gráfico de enjambre de los descriptores del tercer modelo-hito**



Fuente: Elaboración propia

En su interacción, se observa que de dicho resultado menos de un 40% es contribución unitaria, mientras el resto se distribuye principalmente en la interacción con el nivel socioeconómico, la edad de egreso y si egresa o no desde la modalidad científico-humanista.

**Figura 37. Importancia relativa de los descriptores del tercer modelo respecto al género**

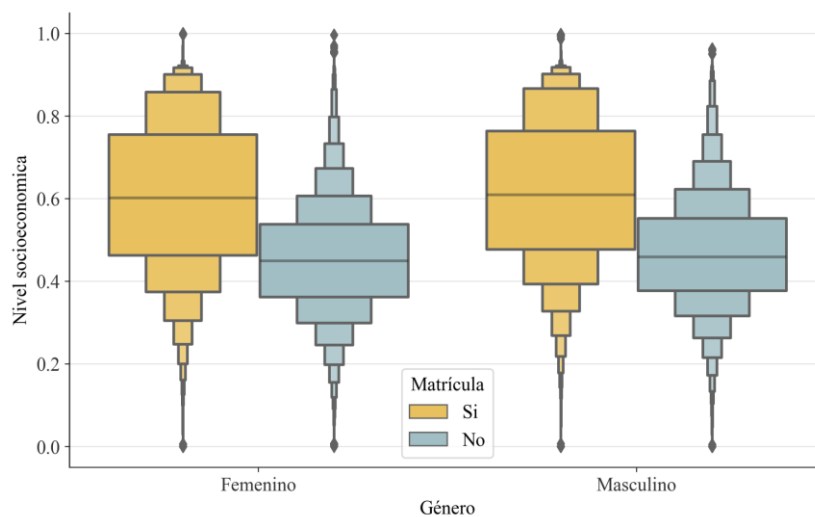


Fuente: Elaboración propia

Sobre esto, considerando que un 50,72% de la población estudiada es de género femenino, el segmento que se matricula en educación superior es femenina en un 52,7% de los casos, lo que visualiza la contribución positiva antes mencionada.

Sobre el nivel socioeconómico, se observa que, si bien existen notorias diferencias en el nivel socioeconómico entre matriculados y aquellos que no, se visualiza también que el efecto de interacción informado por los valores SHAP es insignificante, producto de la baja contribución del predictor de género.

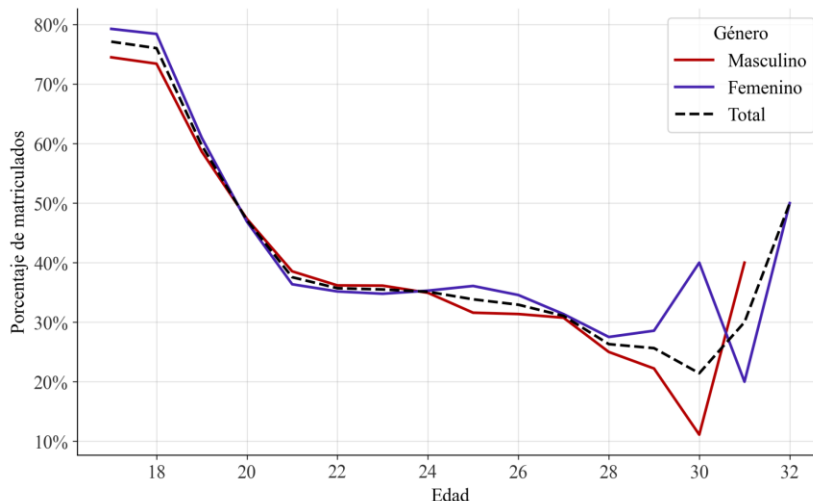
**Figura 38. Distribución de nivel socioeconómico según incidencia de matrícula desglosado por género**



Fuente: Elaboración propia

Luego, sobre la edad de egreso, se observa que en general existe una distancia cercana entre ambos géneros, con un gran porcentaje de matriculados en los primeros años tras su edad ideal de egreso, donde la tasa de matrícula femenina es mayor que la masculina y que ello luego se dinamiza posterior a los 20 años, añadiendo mayor distancia en las tasas de matrícula posterior a los 25 años.

**Figura 39. Tasa de matrícula por edad del egresado**



Fuente: Elaboración propia

Finalmente, se observa que si bien un 51,32% de los egresados desde educación media científico-humanista son de género femenino, este porcentaje es ligeramente mayor al considerar la matrícula, donde el sector de egresados desde dicha forma de educación media desde un 53,18%, lo que es ligeramente superior a la distribución de la población y explica el efecto de interacción antes informado.

Similar al hito de postulación y en línea con lo observado en la revisión de antecedentes sobre el tópico de investigación, se decidió ramificar también el hito de la matrícula para considerar identificar aspectos que potencialmente permitan diferenciar a estudiantes que se matriculan en el área GEEMP de aquellos que no lo hacen y captar la contribución del género en dicha diferenciación.

Para ello seleccionó el modelo entrenado utilizando el algoritmo de *CatBoost* cuyo ajuste, considerando su puntaje de media geométrica, es de 0.641 que es ligeramente inferior a los

modelos anteriores y, además, solo un 14% mejor que el azar, con un detalle de importancia por descriptor que se especifica a continuación.

**Tabla 7. Importancia SHAP de los descriptores en el tercer hito - submodelo GEEMP**

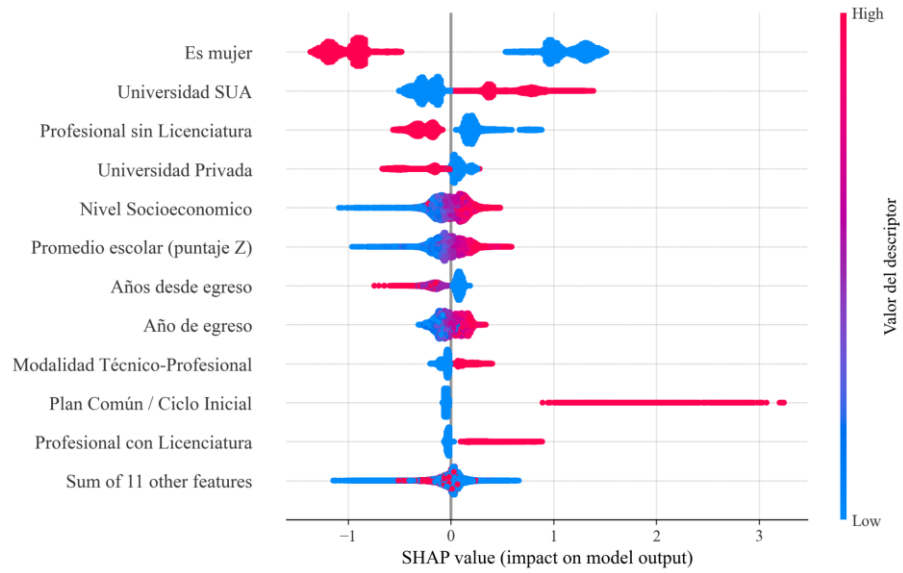
<b>Descriptor</b>	<b>Importancia</b>	<b>Descriptor</b>	<b>Importancia</b>
Es mujer	1,081	Instituto Profesional	0,036
Universidad SUA	0,344	Edad de egreso	0,034
Profesional sin Licenciatura	0,242	Técnico de Nivel Superior	0,023
Universidad Privada	0,131	Licenciatura	0,013
Nivel socioeconómico	0,116	Modalidad Científico-Humanista	0,013
Promedio escolar (puntaje Z)	0,116	Dependencia Part. Pagada	0,012
Años desde egreso	0,110	Dependencia Pública	0,008
Año de egreso	0,103	Centro de Formación Técnica	0,006
Modalidad Técnico-Profesional	0,082	Establecimiento rural	0,006
Plan Común / Ciclo Inicial	0,081	Dependencia Part. Subvencionada	0,006
Profesional con Licenciatura	0,063	Educación adulta	0,005

Fuente: Elaboración propia

Los resultados de dicho modelo nos indican que el descriptor de mayor relevancia es el género, con una relación que indica que pertenecer al género femenino reduce notoriamente las probabilidades de matricularse en el área GEEMP, seguido de aspectos como matricularse en una institución que participe del sistema único de admisión, en una programa de estudios que no sea profesional no conducente a título o en instituciones diferentes a universidades privadas, entre otros descriptores detallados a continuación.



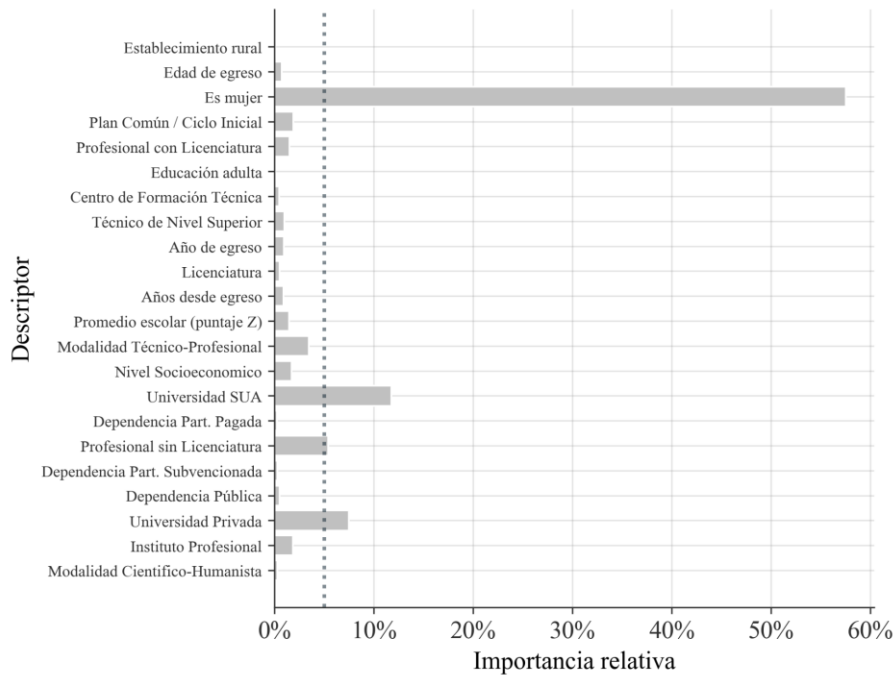
**Figura 40.** Gráfico de enjambre de los valores SHAP del submodelo GEEMP - Tercer hito



Fuente: Elaboración propia

Además de ello, unitariamente el género contribuye casi un 60% por sí solo, seguido de su interacción con matricularse en una institución selectiva vía SUA o en una universidad privada, como se detalla en la siguiente figura.

**Figura 41. Importancia relativa de los descriptores respecto al género**



Fuente: Elaboración propia

En específico, es posible ver qué pese a una matrícula cuya proporción de estudiantes de género femenino es de 52,61%, solo un 20,75% de las estudiantes de matrícula en una carrera del área comentada.

Luego, al evaluar esta situación respecto a si la institución en la que se matriculan es o no selectiva mediante el Sistema Único de Admisión, se observa que mientras la matrícula en estas instituciones es de un 49,57% de estudiantes de género femenino, esta proporción es de un 29,32% en el área GEEMP, lo que es mayor a la tasa global para todas las instituciones, sin embargo, fuera de estas instituciones, la proporción de matrícula por género es de un 53,87% de estudiantes de género femenino, la que se reduce drásticamente en el área GEEMP, con solo un 14,99%.

Finalmente, al considerar a universidades privadas no adscritas al Sistema Único de Admisión, vemos que en ellas la tasa de feminidad es mayor a la distribución global, con un 58,94% de matriculadas, pero, al igual que la tendencia general, dicha proporción se reduce notoriamente en el área GEEMP, con solo un 26,04% de matriculadas.

## 6.5 Retención

A continuación, siguiendo la línea de considerar únicamente a los estudiantes que efectivamente se matriculan en educación superior, se entrenó un modelo para diferenciar a aquellos que, una vez matriculados, presentan retiro inicial o temprano, considerando los 30 descriptores descritos en apartados previos, de los que 27 indican ser significativos.

Para ello consideró al modelo que emplea el algoritmo de *CatBoost*, que obtuvo un resultado muy reducido con una media geométrica de solo 0,578, pero, adicional a ello, los resultados indican que el género tiene una importancia SHAP de solo 0,026.

**Tabla 8. Importancia SHAP de los descriptores en el cuarto hito-modelo**

<b>Descriptor</b>	<b>Importancia</b>	<b>Descriptor</b>	<b>Importancia</b>
Nivel socioeconómico	0,203	Es mujer	0,026
Promedio escolar (puntaje Z)	0,183	Profesional sin Licenciatura	0,026
Beneficiario crédito	0,149	Centro de Formación Técnica	0,020
Técnico de Nivel Superior	0,141	Instituto Profesional	0,016
Educación adulta	0,125	Licenciatura	0,014
Beneficiario beca	0,118	Área LPS	0,013
Área GEEMP	0,117	Dependencia Part. Pagada	0,012
Universidad SUA	0,098	Beneficiario gratuidad	0,011
Postula a BE	0,067	Dependencia. Subvencionada	0,010
Año de egreso	0,061	Profesional con Licenciatura	0,008
Edad de egreso	0,050	Universidad Privada	0,006
Plan Común / Ciclo Inicial	0,041	Otras áreas	0,005
Años desde egreso	0,038	Modalidad Científico-Humanista	0,004
Modalidad Técnico-Profesional	0,033		

Fuente: Elaboración propia

De estos resultados se infiere que el género no es un factor relevante para evaluar si un estudiante se retirará o no de la educación superior de forma inicial o temprana. Sin embargo,

ello no resta la consideración de aspectos de interés en la proporción de estudiantes que desertan por área del conocimiento, en ese sentido se observa que un 42,6% de los estudiantes de género masculino se retira del primer programa de estudios cuando este es GEEMP, respecto a un 39,8% que lo hace cuando es un área diferente. Esto contrasta con las cifras para el género femenino, que indican una tasa de retiro de 45,05% en el área GEEMP respecto a un 33,96% en otras áreas.

## 6.6 Reingreso

Luego, para cada estudiante que se retiró, se elaboró una exploración para indicar diferencias entre el grupo que retorna en los siguientes tres años y aquellos que no presentan reingreso, ello considerando únicamente las transferencias hacia otros programas de estudios.

Para ello se seleccionó el modelo entrenado con el algoritmo de *CatBoost* que obtuvo un desempeño reducido, pero aun así mejor que el modelo descrito en el apartado anterior, con un desempeño de acuerdo con su puntaje de media geométrica de 0,65.

De ello destaca que, similar al modelo anterior, la contribución media del género, su importancia SHAP, es de solo un 0,027, es decir, de acuerdo con el umbral antes definido es una contribución insignificante.

**Tabla 9. Importancia SHAP de los descriptores del quinto modelo-hito**

<b>Descriptor</b>	<b>Importancia</b>	<b>Descriptor</b>	<b>Importancia</b>
Nivel Socioeconómico	0,379	Promedio escolar (puntaje Z)	0,054
Técnico de Nivel Superior	0,355	Beneficiario beca	0,053
Años desde egreso	0,178	Dependencia Part. Pagada	0,051
Beneficiario de crédito	0,098	Dependencia Pública	0,028
Postula a BE	0,096	Es mujer	0,027
Plan Común / Ciclo Inicial	0,093	Profesional sin Licenciatura	0,026
Modalidad Científico-Humanista	0,089	Centro de Formación Técnica	0,026

Descriptor	Importancia	Descriptor	Importancia
Otras áreas	0,084	Universidad Privada	0,027
Profesional con Licenciatura	0,082	Área GEEMP	0,019
Edad de egreso	0,076	Universidad SUA	0,017
Año de egreso	0,071	Licenciatura	0,012
Educación adulta	0,068	Área LPS	0,008
Instituto Profesional	0,062		

Fuente: Elaboración propia

Al igual que en el modelo anterior, pese a que el género no aparece como relevante respecto a otras variables que sí inciden de manera notoria en la diferenciabilidad de estudiantes que reingresan respecto a aquellos que no, si es relevante hacer notar aspectos como que, si bien la proporción por género de estudiantes que se retiraron de su primer programa de estudios es de un 47,89% femenina, al evaluarlo respecto a los que retornan y por área desde la que se retiraron, observamos que en el segmento masculino la tasa de retorno es de más o menos un 65% indiferente del área, cifra similar a la tasa de retorno de las estudiantes que retornan después de desertar en áreas diferentes a GEEMP, sin embargo, la tasa de retorno de estudiantes de género femenino que se retiró de un programa de estudios en el área GEEMP es de un 79,94%.

En función de ello, se visualiza a continuación los resultados del modelo específico para identificar diferencias en estudiantes que, presentando un retiro anterior, reingresan a un programa de estudios diferente y si este es o no del área GEEMP. Para ello, empleando el modelo entrenado con de *Histogram Based Gradient Boosted Classifier* con un puntaje de media geométrica de 0,712, se determinó que retirarse desde el área GEEMP es notoriamente el mayor predictor del reingreso hacia esta área, seguida por el género del estudiante y si el retiro ocurrió o no en una institución participante del Sistema Único de Admisión.

**Tabla 10. Importancia SHAP de los descriptores del quinto hito - submodelo GEEMP**

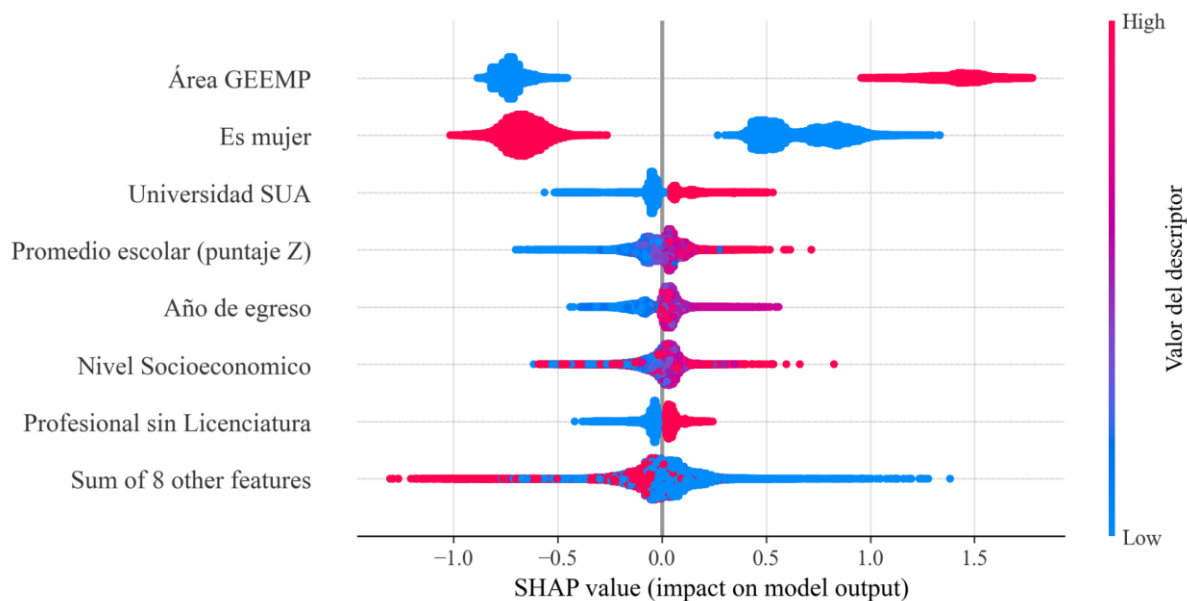
Descriptor	Importancia	Descriptor	Importancia
Área GEEMP	0,957	Dependencia Part. Pagada	0,033

Descriptor	Importancia	Descriptor	Importancia
Es mujer	0,665	Modalidad Científico-Humanista	0,032
Universidad SUA	0,071	Beneficiario beca	0,024
Promedio escolar (puntaje Z)	0,068	Plan Común / Ciclo Inicial	0,023
Año de egreso	0,063	Técnico de Nivel Superior	0,022
Nivel socioeconómico	0,055	Instituto Profesional	0,017
Profesional sin Licenciatura	0,053	Área LPS	0,016
Universidad Privada	0,033		

Fuente: Elaboración propia

En lo específico, a la vez que retirarse desde el área GEEMP contribuye positivamente hacia el reingreso en esta área, que el género del estudiante sea femenino implica también una reducción en las posibilidades de reingreso hacia esta misma área.

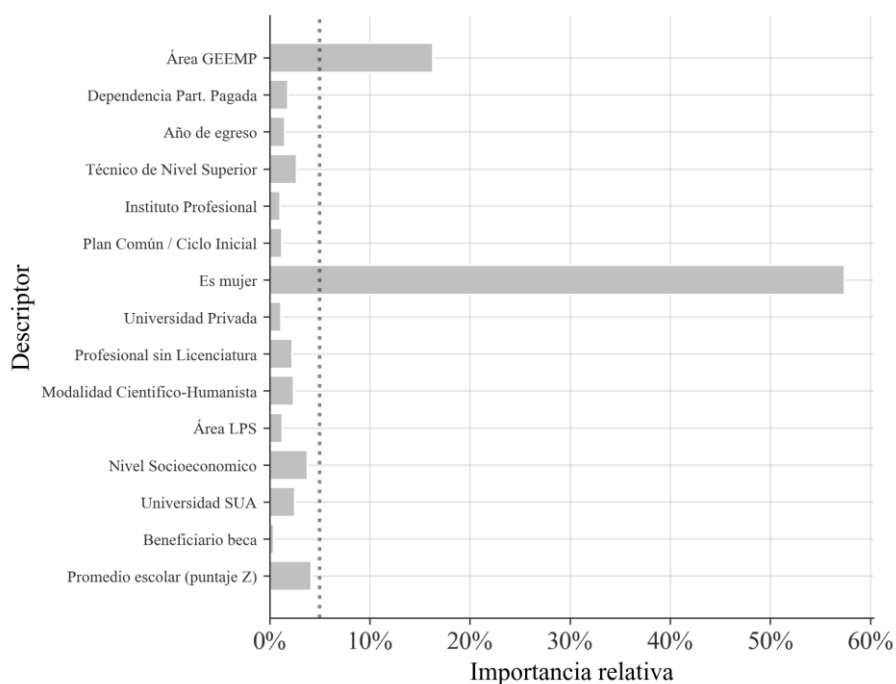
Figura 42. Valores SHAP del submodelo GEEMP - Quinto hito



Fuente: Elaboración propia

En su interacción, el género contribuye unitariamente el 58% de su propio aporte, con su efecto de interacción más importante y el único superior al umbral de relevancia del 5% de la contribución total.

**Figura 43. Importancia relativa de los descriptores respecto al género en el submodelo GEEMP - Quinto hito**



Fuente: Elaboración propia

En específico, esto implica que la tasa transición retiro-reingreso para las estudiantes, entendida ésta como el porcentaje de retiros de un área que reingresa hacia un área igual o diferente, es como se describe en la siguiente tabla.

**Tabla 11. Transito porcentual desde área de retiro y área de reingreso en el género femenino - Porcentaje por área de retiro**

Área de retiro	Área de reingreso		
	GEEMP	LPS	Otras áreas
<b>GEEMP</b>	32,5%	29,2%	38,2%
<b>LPS</b>	0,05%	60,5%	34,6%
<b>Otras áreas</b>	0,05%	22,3%	72,5%

Fuente: Elaboración propia

Mientras que, por su lado, dicha tasa de transferencia para los estudiantes de género masculino es como se exhibe a continuación.

**Tabla 12. Transito porcentual desde área de retiro y área de reingreso en el género masculino - Porcentaje por área de retiro**

Área de retiro	Área de reingreso		
	GEEMP	LPS	Otras áreas
<b>GEEMP</b>	59,7%	10,1%	30,2%
<b>LPS</b>	18,6%	45,3%	36,1%
<b>Otras áreas</b>	20,0%	13,3%	66,7%

Fuente: Elaboración propia

De lo anterior es posible comprender que, primero, el aspecto de elección de carrera diferenciada por género permanece, con una transferencia desde otras áreas hacia GEEMP que, en el género masculino, es de alrededor del 20%, pero para las estudiantes ronda el 5%, sin embargo, lo más destacable es que la tasa transferencia GEEMP-GEEMP en las estudiantes es de alrededor la mitad que en el género contrario, cuya permanencia en el mismo área GEEMP es del 59,7%, ello en contraste con la transferencia LPS-LPS y desde otras áreas hacia otras áreas, que exhiben una mayor permanencia intra área en las estudiantes que en el género opuesto.



## 7 Discusión

En general, las trayectorias en educación terciaria para la población estudiada, esto es, aquellas cuyo egreso ocurrió entre 2008 y 2014, ambas cohortes inclusive, y que además llevan al menos 4 años de matrícula escolar desde el primer período de información disponible, registran una notoria desigualdad vinculada, principalmente, a la elección de carrera, que les resta notoriamente de las áreas del conocimiento propias del segmento GEEMP (del inglés Geociencias, ingeniería, economía, matemáticas/ciencias de la computación y ciencias físicas) y las sobrerrepresenta ligeramente en LPS (en inglés, Ciencias de la vida, psicología y ciencias sociales). Además, un segundo aspecto relevante es su menor reingreso, posterior al retiro desde su primera matrícula, hacia áreas GEEMP.

Esto indica que no solo existe un aspecto vocacional y de elección de carrera o, en paralelo, de selección de estudiantes, sino además de retención de estas, sugiriendo la posible existencia de factores de expulsión de las estudiantes como es sugerido en la literatura, aunque desafortunadamente no se cuenta con información sistematizada para identificar dichos aspectos.

Adicional a ello, es remarcable que en el egreso escolar las estudiantes presentan un desempeño escolar ligeramente superior a sus contrapartes masculinas, junto con un ligeramente mayor nivel socioeconómico y un egreso más frecuente desde la enseñanza media Científico-Humanista a la vez que menor egreso desde la modalidad para jóvenes y adultos, sin embargo, considerando el desempeño del clasificador entrenado de solo 0,578, podemos asumir que en su egreso ambos grupos son notoriamente homogéneos.

De la interacción entre pares de variables, podemos observar cómo existe un leve patrón de correlación entre el nivel socioeconómico, el género y el tipo de enseñanza media de egreso, de modo que las mujeres que egresan de dicho tipo de educación exhiben un nivel socioeconómico ligeramente mayor.

A la vez, respecto al promedio escolar, se observa una pequeña sobrerrepresentación femenina en los segmentos inferiores de nivel socioeconómico que, a la vez, exhiben un menor promedio escolar, lo que indica que un mayor desempeño escolar femenino es mucho más frecuente en los segmentos de menor ingreso.

Un último aspecto por destacar es una ligera tendencia hacia un mejor nivel socioeconómico en los egresos desde la modalidad PJA femeninos respecto a los masculinos, sin embargo, del mismo modo que los dos elementos antes destacados, a pesar de existir interacción entre las variables, ella resulta prácticamente insignificante frente a la mucho más notoria y ya mencionada homogeneidad entre ambos grupos.

Luego, sobre la postulación mediante el Sistema de Admisión Universitaria, podemos observar que, en el tiempo, la postulación femenina ha aumentado ligeramente su distancia respecto a la masculina, pasando desde tener una tasa de postulantes de un 34%, con una distancia de 2 puntos porcentuales respecto a la masculina en la cohorte de egreso del año 2008, hasta una tasa del 44%, con una distancia de 3 puntos respecto a la masculina en la última cohorte trazada.

Además de ello, se visualizan dos relaciones débiles entre postular al Sistema Único de Admisión, un mejor desempeño escolar y el género femenino y la misma situación en cuanto al nivel socioeconómico, pero ambos grupos a pesar de presentar diferencias estadísticas, ellas no son de magnitud suficiente como para evaluarlas como de interés en la distinción entre grupos.

En cuanto a la postulación al interior del área GEEMP, para los estudiantes de la población estudiada que efectivamente postulan mediante el SUA, se observa que, luego del puntaje en matemáticas, el género es el segundo descriptor más relevante para diferenciar quienes postulan a dicha área en sus tres primeras preferencias.

Al respecto, a pesar de presentarse efectos de interacción con el puntaje de ambas pruebas específicas clasificados como relevantes de acuerdo con el umbral seleccionado, ninguna de dichas diferencias se reporta como relevante para la distinción entre ambos grupos, aunque es menester observar que las estudiantes reportan un mejor desempeño en la prueba específica de lenguaje, mientras los estudiantes lo hacen en la de matemáticas.

Siguiendo con la matrícula, se observa que el género es un descriptor de poca relevancia en la diferenciación de qué estudiantes se matriculan en la educación terciaria respecto a aquellos que no, con una interacción respecto a la edad de egreso, el nivel socioeconómico y el tipo de enseñanza media.

Sin embargo, pese a que dichos efectos de interacción se pueden observar en una ligeramente mayor matrícula femenina posterior a los 25 años y dos puntos porcentuales de diferencia respecto al egreso masculino en los egresos de enseñanza media Científico-Humanista, ninguna de dichas interacciones es suficientemente relevante para generar diferenciación notoria, ello producto de la baja relevancia antes mencionada al género como descriptor en el evento estudiado.

Indiferente de ello, al considerar únicamente a los matriculados, se observa que la importancia del género como descriptor es mucho mayor respecto a otros elementos, con un aporte unitario que ronda el 58% de su importancia y efectos de interacción sobre la matrícula en instituciones universitarias de ambos tipos (participantes del SUA y aquellas que no). Ambos efectos de interacción indican que, si bien el género por sí solo es un factor de diferenciación relevante, esta brecha es menor en instituciones universitarias, con una diferencia de 9 puntos porcentuales en instituciones selectivas y 6 respecto para las no participantes del SUA.

En cuanto a la retención, se observa que el género por sí solo ni en interacción con otros descriptores es un factor relevante de diferenciación sobre qué estudiantes se retiran de manera inicial o temprana de sus programas de estudio, ello indica que, si bien pueden existir diferencias por género en la retención para cada área, otros descriptores sobre los que se controla asimilan dicha diferenciación y esta es, por tanto, más relevante que el género por sí solo.

Una situación similar ocurre para el reingreso, donde el género no es un factor de relevancia, sin embargo, al desglosar dicho hito en reingreso hacia el área GEEMP respecto a aquellos que no, visualizamos dos aspectos de interés.

Primero, que existe un menor reingreso hacia GEEMP de aquellas estudiantes que se retiran desde dicha área y, además, que el retiro desde otras áreas también exporta una menor cantidad de estudiantes hacia GEEMP. Ambas situaciones indican un asunto vocacional, de elección de carrera para las estudiantes, pero también un aspecto de retención que ya fue mencionado inicialmente, es decir, no solo las estudiantes optan menos por dicha área de estudios, sino que además si lo hacen en su primera matrícula y se retiran desde esta, su reingreso es más frecuentemente en otras áreas que en la misma.

Lo anterior, desafortunadamente, no expresa efectos de interacción relevantes que permitan identificar subgrupos en la incidencia del fenómeno comentado.

## **7.1 Limitaciones**

El aporte de emplear datos cuasi-censales y administrativos es también una notoria limitación en la consideración de las trayectorias, pues a la vez que permiten trazar completamente el paso por la educación terciaria de los estudiantes, omiten la posibilidad de visualizar aspectos sustantivos al interior del aula, ya sean estos de carácter individual como la decisión de suspender estudios institucionalmente, el desempeño del estudiante, la cantidad de materias por semestre y año o rezago por reprobación de asignaturas. Así como tampoco permiten visualizar aspectos interpersonales, como el clima al interior del programa de estudios, la relación con las autoridades académicas, formas de discriminación existentes, entre otras.

Además de lo anterior, la presente aproximación está limitada temporalmente, puesto que se propone conocer a la vez los últimos cuatro años de educación escolar del estudiante, como también los ocho años posteriores, ello considerando que se dispone de información para únicamente 14 períodos, significa que las cohortes trazables no consideran a los ingresos más recientes, omitiendo el aporte que podrían realizar elementos que se han acrecentado en los años recientes, ya sea las formas de estudio durante la pandemia, las manifestaciones universitarias en 2018 o el aumento de cobertura de la gratuidad.

Adicional a ello, existen severas limitaciones en el estudio de trayectorias en el nivel terciario, marcadas principalmente la heterogeneidad existente en ellas, los tiempos disímiles de ocurrencia de los diferentes hitos aún dentro de una misma cohorte y como dichas dificultades se incrementan conforme aumentan también los períodos desde el ingreso, siendo posible y difícil de sistematizar aspectos como el reingreso al mismo programa de estudios, el tránsito hacia menciones o desde ciclos iniciales, la articulación de estudios, entre otros.

## 8 Conclusiones

Es posible confirmar que existen diferencias entre estudiantes de género femenino y masculino durante el egreso, con una mayor tasa de egreso desde enseñanza media Científico-Humanista entre las estudiantes, junto a un proporcionalmente menor egreso desde educación para jóvenes y adultos y un mejor desempeño escolar, sin embargo, en términos generales ambos grupos son notoriamente homogéneo y ninguna de las tres diferencias es de gran relevancia.

Adicionalmente, se corrobora que, en la postulación, si bien las estudiantes tienen una mayor tasa de participación en el Sistema Único de Admisión, se confirma la hipótesis expuesta sobre una menor orientación hacia las áreas de mayor masculinización (específicamente GEEMP), además de identificar que también están más predispuestas a postular hacia carreras feminizadas como son LPS.

Sobre la matrícula, se identifica que la homogeneidad entre ambos segmentos es notoriamente alta, aunque también, al igual que con la hipótesis previa, se corrobora una situación análoga de mayor orientación vocacional hacia LPS y menor hacia GEEMP. Adicionalmente a ello, dicha desigualdad es menor en instituciones universitarias, participantes o no del Sistema Único de Admisión.

Luego, se identifica que indiferente del género y/o el área de estudios, la retención es homogénea, siendo otros aspectos mucho más decisivos en su consideración.

Y, finalmente, en la transición entre programas de estudio posteriores a un retiro, se observa que existe, primero, una menor transición desde otras áreas, pero también desde la misma, ello explicado principalmente por el género en sí mismo sin interacción con otros descriptores disponibles.

En cuanto a la hipótesis de interacción entre factores de desigualdad, no se encuentra evidencia para indicar que existan subgrupos con mayor desigualdad, a excepción de la ya comentada situación con la matrícula universitaria del área GEEMP.

Esto último es indicativo de dos escenarios posibles: Primero, existen aspectos ya comentados en las limitaciones sobre los que no se dispone información producto del carácter

administrativo de los datos disponibles y que podrían delimitar líneas de investigación futuras; segundo, que existan factores de desigualdad de mucha mayor relevancia, como el nivel socioeconómico, según sugieren los resultados, y cuya interacción con el género sea reducida y, por tanto, sea así el género un aspecto mucho menos relevante en las trayectorias fuera del apartado vocacional y de retención por área. No obstante, este tipo de investigación cuantitativa levanta la necesidad de estudiar cualitativamente los fenómenos de la elección de carreras GEEMP y que pasa en el interior de las instituciones

Se comenta además la poca relevancia observada a aspectos como la dependencia de egreso o la comuna de residencia del estudiante, ello producto, probablemente, de su correlación con aspectos como el nivel socioeconómico, de modo que estas pasaron a segundo plano respecto a una variable que contribuye mayor información a los diferentes modelos.

Eventuales futuras líneas de investigación podrían nutrirse de incorporar aspectos territoriales o institucionales que no sean captados por ambas variables, como la migración hacia las sedes de educación superior, el desempeño de los establecimientos escolares de egreso o la transición entre ramas específicas de la educación media Técnico-Profesional y que no fueron abordadas en la presente investigación.

Además, se hace evidente la limitación del uso de datos administrativos y la necesidad de explorar factores intra-aula e interpersonales al interior de las instituciones y los programas educativos.

## Referencias

- Baeza, A., & Lamadrid, S. (2018). Trayectorias educativas según género. Lo invisible para la política educativa chilena. *Revista de Investigación Educativa*, 36(2), Art. 2.  
<https://doi.org/10.6018/rie.36.2.298061>
- Bentéjac, C., Csörgő, A., & Martínez-Muñoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3), 1937-1967.  
<https://doi.org/10.1007/s10462-020-09896-5>
- Blanco, C., Meneses, F., & Paredes, R. (2018). Más allá de la deserción: Trayectorias académicas en la educación superior en Chile. *Calidad en la Educación*, 49, Art. 49.  
<https://doi.org/10.31619/caledu.n49.579>
- Bordón, P., Canals, C., & Mizala, A. (2020). The gender gap in college major choice in Chile. *Economics of Education Review*, 77, 102011.  
<https://doi.org/10.1016/j.econedurev.2020.102011>
- Browne, M. (2017). Análisis del Sistema de Administración Delegada creada por el DL N° 3166 de 1980. *Ministerio de Educación-SETP*.  
<http://biblioteca.digital.gob.cl/handle/123456789/897>
- Carrasco, E. (2016). *Elección de carrera en estudiantes de nivel socioeconómico bajo en Chile: Variables asociadas, tensiones y desafíos*.  
<https://repositorio.uchile.cl/handle/2250/153745>

Castillo, R., Grazzi, M., & Tacsir, E. (2014). *Women in Science and Technology. What does the literature say?* (IDB-TEN-637; Número IDB-TEN-637). Inter-American Development Bank.

Castillo-Peña, J. (2021). Expectativas y trayectorias educativas postsecundarias de jóvenes de territorios rurales en Chile. Una mirada desde el desarrollo humano. *Revista iberoamericana de educación superior*, XII(34), Art. 34.

Charlesworth, T. E. S., & Banaji, M. R. (2019). Gender in Science, Technology, Engineering, and Mathematics: Issues, Causes, Solutions. *Journal of Neuroscience*, 39(37), Art. 37.  
<https://doi.org/10.1523/JNEUROSCI.0475-18.2019>

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin*, 143, 1-35.  
<https://doi.org/10.1037/bul0000052>

CPCE. (2020). Enfoque de políticas ESTP – N°10: Financiamiento comparado de la educación superior en Chile con foco en la educación superior técnico profesional. *Centro de Políticas Comparadas de Educación UDP - Universidad Diego Portales*.  
<https://cpce.udp.cl/publicacion/enfoque-de-politicas-estp-no10-financiamiento-comparado-de-la-educacion-superior-en-chile-con-foco-en-la-educacion-superior-tecnico-profesional/>

Delisle, J., & Bernasconi, A. (2018). *Lessons from Chile's transition to free college* (N.º 43; Evidence speaks reports, Número 43). Economic Studies at Brookings.  
<https://www.brookings.edu/research/lessons-from-chiles-transition-to-free-college/>



Dhamdhere, K., Agarwal, A., & Sundararajan, M. (2020). The Shapley Taylor interaction index. *Proceedings of the 37th International Conference on Machine Learning*, 9259-9268.

Diaz, O. E., Guajardo, D. C., & Fiegehen, L. E. G. (2017). Educación de adultos en Chile: Percepciones y valoraciones de sus beneficiarios. *Revista Brasileira de Educação*, 22. <https://doi.org/10.1590/S1413-24782017227176>

Donoso, S., & Schiefelbein, E. (2007). Analysis of explicative models of student' s retention at the University: A vision from the social inequality. *Estudios pedagógicos (Valdivia)*, 33(1), Art. 1. <https://doi.org/10.4067/S0718-07052007000100001>

Fincham, E., Rózemberczki, B., Kovanović, V., Joksimović, S., Jovanović, J., & Gašević, D. (2021). Persistence and Performance in Co-Enrollment Network Embeddings: An Empirical Validation of Tinto's Student Integration Model. *IEEE Transactions on Learning Technologies*, 14(1), Art. 1. <https://doi.org/10.1109/TLT.2021.3059362>

Gaete, M., & Morales, R. (2011). Articulación del sistema de educación superior en Chile: Posibilidades. tensiones y desafíos. *Calidad en la Educación*, 35, Art. 35. <https://doi.org/10.31619/caledu.n35.96>

Ganley, C. M., George, C. E., Cimpian, J. R., & Makowski, M. B. (2018). Gender Equity in College Majors: Looking Beyond the STEM/Non-STEM Dichotomy for Answers Regarding Female Participation. *American Educational Research Journal*, 55(3), Art. 3. <https://doi.org/10.3102/0002831217740221>

- Guevara, E., & García, A. (2010). Orden de género y trayectoria escolar en mujeres estudiantes de ciencias exactas y naturales. *Investigación y Ciencia*.  
<https://www.redalyc.org/articulo.oa?id=67413508003>
- Hägglund, A. E., & Leuze, K. (2021). Gender differences in STEM expectations across countries: How perceived labor market structures shape adolescents' preferences. *Journal of Youth Studies*, 24(5), Art. 5. <https://doi.org/10.1080/13676261.2020.1755029>
- Huang, J., Gates, A. J., Sinatra, R., & Barabási, A.-L. (2020). Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences*, 117(9), Art. 9. <https://doi.org/10.1073/pnas.1914221117>
- Kursa, M. B., Jankowski, A., & Rudnicki, W. R. (2010). Boruta – A System for Feature Selection. *Fundamenta Informaticae*, 101(4), 271-285. <https://doi.org/10.3233/FI-2010-288>
- Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2019). Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv:1802.03888 [cs, stat]*. <http://arxiv.org/abs/1802.03888>
- Lundberg, S. M., Nair, B., Vavilala, M. S., Horibe, M., Eisses, M. J., Adams, T., Liston, D. E., Low, D. K.-W., Newman, S.-F., Kim, J., & Lee, S.-I. (2018). Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, 2(10), 749-760. <https://doi.org/10.1038/s41551-018-0304-0>
- Milo, T., & Somech, A. (2020). Automating Exploratory Data Analysis via Machine Learning: An Overview. *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, 2617-2622. <https://doi.org/10.1145/3318464.3383126>

Moss-Racusin, C. A., Sanzari, C., Caluori, N., & Rabasco, H. (2018). Gender Bias Produces Gender Gaps in STEM Engagement. *Sex Roles*, 79(11), Art. 11.

<https://doi.org/10.1007/s11199-018-0902-z>

Obregón, V., & Sepúlveda, P. (2019). *Pasar del dicho al hecho. Desafíos en la Educación Superior*. <http://repositorioslatinoamericanos.uchile.cl/handle/2250/3056855>

Pareja, C., Mac-Clure, O., & Pérez, C. (2021). Acceso a la educación universitaria y gratuidad: Movilidad educacional y movilidad territorial en una región no metropolitana de Chile. *Calidad en la Educación*, 55, Art. 55. <https://doi.org/10.31619/caledu.n55.1021>

Peña y Lillo, A. y L. (2006). *Elección de carreras universitarias y género. El caso del Trabajo Social e Ingeniería*. <https://repositorio.uchile.cl/handle/2250/113536>

Ramos, C. F. (2018). *Decisiones educativas y valoración de la educación superior en la trayectoria académica de los profesionales primera generación universitaria*.

<https://repositorio.uchile.cl/handle/2250/168495>

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). «Why Should I Trust You?»: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144.

<https://doi.org/10.1145/2939672.2939778>

Richardson, S. S., Reiches, M. W., Bruch, J., Boulicault, M., Noll, N. E., & Shattuck-Heidorn, H. (2020). Is There a Gender-Equality Paradox in Science, Technology, Engineering, and Math (STEM)? Commentary on the Study by Stoet and Geary (2018). *Psychological Science*, 31(3), Art. 3. <https://doi.org/10.1177/0956797619872762>

- Rodríguez, P., & Villanueva, A. (2022). Design, Development, and Evaluation of a Predictive Model for Regular School Dropout in the Chilean Educational System. En S. Hosseini, D. H. Peluffo, J. Nganji, & A. Arrona-Palacios (Eds.), *Technology-Enabled Innovations in Education* (pp. 493-505). Springer Nature. [https://doi.org/10.1007/978-981-19-3383-7\\_40](https://doi.org/10.1007/978-981-19-3383-7_40)
- Sepúlveda, M. J., & Manquepillán, M. (2017). Doc. de trabajo N° 2: Brechas de género en el Sistema Único de Admisión a la educación superior. *Ministerio de Educación-Centro de Estudios*. <http://biblioteca.digital.gob.cl/handle/123456789/390>
- Stewart-Williams, S., & Halsey, L. G. (2021). Men, women and STEM: Why the differences and what should be done? *European Journal of Personality*, 35(1), Art. 1. <https://doi.org/10.1177/0890207020962326>
- Stoet, G., & Geary, D. C. (2018). The Gender-Equality Paradox in Science, Technology, Engineering, and Mathematics Education. *Psychological Science*, 29(4), Art. 4. <https://doi.org/10.1177/0956797617741719>
- Thébaud, S., & Charles, M. (2018). Segregation, Stereotypes, and STEM. *Social Sciences*, 7(7), Art. 7. <https://doi.org/10.3390/socsci7070111>
- Venegas-Muggli, J. I. (2022). Secondary School Context and (Timely) College Completion in Chile's Education System: A Cohort Study. *Journal of College Student Retention: Research, Theory and Practice*, 24(1), Art. 1. Scopus. <https://doi.org/10.1177/1521025120910477>

Yáñez, K. G. D., Tropa, J. A. R., & Quilamán, J. P. Q. (2019). Brechas de género en los resultados de pruebas de selección universitaria en Chile. ¿Qué sucede en los extremos superior e inferior de la distribución de puntajes? *Pensamiento Educativo, Revista de Investigación Latinoamericana (PEL)*, 56(1), Art. 1.  
<https://doi.org/10.7764/10.7764/PEL.56.1.2019.5>

## Tablas

**Tabla 13. Descripción de beneficios estatales para el financiamiento de la educación terciaria**

Nombre	Año de inicio	Beneficiarios	Observaciones	Cobertura
Becas de arancel				
Beca de Apoyo Norte Valparaíso (BANV)	2017	Estudiantes de educación superior cuyo grupo familiar haya resultado gravemente afectado por el alud ocurrido en las regiones Antofagasta y Atacama en el año 2015. Siempre que para el año 2017 cuenten con la calidad de alumno regular en una institución de educación superior reconocida por el Estado.	Concedida por JUNAEB. Considera también un beneficio para alimentación y otro para mantención.	Arancel referencial de la carrera fijado por MINEDUC.
Beca Universidad ARCIS (BARCIS)	2018	Estudiantes pertenecientes al 70% de menores ingresos, matriculados en la Universidad ARCIS desde 2015 en adelante, que aún no hayan egresado de sus carreras y se matriculen durante 2018 en instituciones de Educación Superior acreditadas.	Beca de reubicación y continuidad dado el cierre de la Universidad ARCIS iniciado el año 2016.	Arancel referencial de la carrera fijado por MINEDUC.
Beca de Articulación (BAR)	s/i	Estudiantes del 70% de menores ingresos, egresados o titulados de carreras técnicas de nivel superior durante los dos años anteriores que quieran continuar sus estudios para obtener un título profesional, en un área del conocimiento afín a su carrera de origen y en una institución acreditada.		Hasta \$750.000 del arancel anual de la carrera.
Beca Bicentenario (BBIC)	s/i	Alumnos pertenecientes al 70% de menores ingresos, egresados de enseñanza media con rendimiento académico meritorio, matriculados en universidades pertenecientes al CRUCH o privadas, acreditadas por al menos cuatro años.	Desde 2021, esta beca cuenta con cupos especiales para estudiantes en situación de discapacidad.	Arancel referencial de la carrera fijado por MINEDUC.
Beca de Excelencia Académica (BEA)	s/i	Alumnos pertenecientes al 80% de menores ingresos que ingresan al año siguiente de su egreso de educación secundaria a la educación superior a una institución acreditada, con un promedio de notas de enseñanza media dentro del 10% más alto del establecimiento y provenientes de un establecimiento de dependencia municipal, particular subvencionado o de administración delegada.		Hasta \$1.150.000 del arancel anual de la carrera.
Beca de Excelencia Técnica (BET)	s/i	Alumnos pertenecientes al 70% de menores ingresos que ingresan dentro de los cuatro años siguientes a su egreso de educación secundaria a la educación superior a una carrera técnica impartida por institución acreditada, con un promedio de notas de enseñanza media superior a 5,0 y con una selección basada en estricto orden relativo al puntaje NEM hasta suplir cupos disponibles.	Corresponde a cupos especiales dentro de la Beca Nuevo Milenio	Hasta \$900.000 del arancel anual de la carrera.

Nombre	Año de inicio	Beneficiarios	Observaciones	Cobertura
Beca Universidad Iberoamericana (BIBERO)	2018	Estudiantes pertenecientes al 70% de menores ingresos, matriculados en la Universidad Iberoamericana desde 2017 en adelante, que aún no hayan egresado de sus carreras y se matriculen durante 2018 en instituciones de Educación Superior acreditadas.	Beca de reubicación y continuidad dado el cierre de la Universidad Iberoamericana iniciado el año 2017.	Arancel referencial de la carrera fijado por MINEDUC.
Beca Hijo de Profesionales de la Educación (BHPE)	s/i	Alumnos del 80% de menores ingresos cuyos padres sean profesores o asistentes de la Educación de enseñanza básica o media en establecimientos municipales, particulares subvencionados o de administración delegada y se matriculen en una institución acreditada.		Hasta \$500.000 del arancel anual de la carrera.
Beca Juan Gómez Millas (BJGM)	s/i	Estudiantes con rendimiento académico meritório, pertenecientes al 70% de menores ingresos del país y matriculados en una institución acreditada.	Esta beca cuenta con cupos especiales para estudiantes en situación de discapacidad.	Hasta \$1.150.000 del arancel anual de la carrera.
Beca Juan Gómez Millas para Extranjeros (BJGME)	s/i	Estudiantes extranjeros con permanencia definitiva o con residencia, que hayan cursado la educación secundaria en Chile, pertenezcan al 70% de menores ingresos y estén matriculados en una institución acreditada.		Hasta \$1.150.000 del arancel anual de la carrera.
Beca Nuevo Milenio (BNM; BNM I; BNM II; BNM III)	s/i	Estudiantes que pertenezcan al 70% de menores ingresos y estén matriculados en una institución acreditada.	Los tramos de cobertura dependen del nivel de ingresos del estudiante.	Hasta \$860.000 del arancel anual de la carrera.
Beca de Pedagogía (BPED)	s/i	Estudiantes matriculados por primera vez y en primer año en carreras de pedagogía, educación parvularia o educación diferencial en instituciones acreditadas y elegibles para este beneficio.	Una vez obtenido el título profesional, el estudiante becado debe trabajar por al menos 3 años en establecimientos educacionales dependientes de los Servicios Locales de Educación Pública, Municipales, Particulares Subvencionados o de Administración Delegada, donde tendrá que cumplir una jornada laboral docente de 44 horas semanales o su equivalente.	Totalidad del arancel anual de la carrera, más un aporte mensual de \$80.000.
Beca de Nivelación Académica (BNA)	2011	Beca asignada a los estudiantes que resulten seleccionados en instituciones con un programa de nivelación académica aprobado por MINEDUC.		
Beca Cumplimiento de Sentencia y Acuerdos (BNC)		Victimas vinculadas a casos específicos de violación a derechos fundamentales bajo responsabilidad del Estado.	Considera los casos Norín Catrimán y Lemún Saavedra.	

Nombre	Año de inicio	Beneficiarios	Observaciones	Cobertura
Beca de Continuidad Universidad del Pacífico (BPACIFICO)	2019	Estudiantes pertenecientes al 70% de menores ingresos, matriculados en la Universidad Iberoamericana desde 2018 en adelante, que aún no hayan egresado de sus carreras y se matriculen durante 2019 en instituciones de Educación Superior acreditadas.	Beca de reubicación y continuidad dado el cierre de la Universidad Iberoamericana iniciado el año 2018.	Arancel referencial de la carrera fijado por MINEDUC.
Beca Puntaje PSU (BPSU)	s/i	Estudiantes pertenecientes al segmento de 80% de menores ingresos y que obtengan puntaje nacional en cualquiera de las pruebas PSU o como promedio de las pruebas de Lenguaje y Matemáticas, el mismo año en que se postula a la beca; y que ingresen a la Educación Superior al año siguiente de egresar de Enseñanza Media. Deben provenir de colegios municipales, particulares subvencionados o de administración delegada.	Desde el proceso de admisión 2020, existe una beca similar llamada Puntaje PDT y relativa al puntaje obtenido en esta.	Hasta \$1.150.000 del arancel anual de la carrera.
Becas de Reparación (RETTIG; TITULAR; TRASPASO)	s/i	Estudiantes de pregrado cuyos padres, madres, abuelos o abuelas estén en la nómina de informe Valech o RETTIG de víctimas de prisión o tortura por parte del Estado durante dictadura militar.	Los titulares de la beca pueden traspasar el beneficio a un descendiente (hasta el segundo grado de consanguinidad, es decir, hijos y nietos).	Totalidad de la matrícula y arancel anual de la carrera. En caso de traspaso, hasta \$1.150.000 del arancel anual de la carrera o \$600.000 en caso de carreras técnicas.
Beca Universidad del Mar (REUBICACION)	2013	Estudiantes matriculados en la Universidad del Mar el año 2012 y que se matriculen en una institución acreditada.		Arancel referencial de la carrera fijado por MINEDUC.
<b>Créditos</b>				
Crédito con Garantía Estatal	s/i	Estudiantes de pregrado que requieran apoyo para financiar sus estudios superiores.	Desde 2014 no hay requisitos socioeconómicos.	Arancel referencial de la carrera fijado por MINEDUC.
Fondo Solidario de Crédito Universitario (FSCU)	s/i	Estudiantes de pregrado, pertenecientes al 80% de menores ingresos, con un puntaje PSU mayor a 475 (promedio Lenguaje y Matemáticas), que requieran apoyo para financiar sus estudios superiores y se matriculen en una universidad del CRUCH acreditada.		Arancel referencial de la carrera fijado por MINEDUC para estudiantes pertenecientes al 60% de menores ingresos. Para el tramo entre 60% y 80%, se financia entre un 50% y un 100% del arancel.
<b>Gratuidad</b>				
Gratuidad	2016	Estudiantes que pertenecen al 60% de menores ingresos del país y estén matriculados en una institución adscrita al beneficio.	Sin requisitos de renovación, se otorga para el tiempo establecido de duración de la carrera.	Arancel y matrícula anual de la carrera.



Fuente: Elaboración propia



Descriptor	Modelo / Hito							
	Egreso	Postulación	Postulación (GEEMP)	Matrícula	Matrícula (GEEMP)	Retención	Reingreso	Reingreso (GEEMP)
Años desde el egreso			X		X	X	X	X

Fuente: Elaboración propia

**Tabla 15. Categorización STEM por clasificación CINE**

Clasificación CINE - 2013		Categoría STEM
Área	Subárea	
Administración de Empresas y Derecho	Derecho	No STEM
	Educación Comercial y Administración	
Agricultura, Silvicultura, Pesca y Veterinaria	Silvicultura	LPS
	Veterinaria	
	Agricultura	
	Pesca	
	Agricultura, Silvicultura, Pesca y Veterinaria sin mayor definición	
Artes y Humanidades	Artes	No STEM
	Lenguajes	
	Idiomas	
	Humanidades	
	Artes y Humanidades sin mayor definición	
Ciencias Sociales, Periodismo e Información	Ciencias Sociales y del Comportamiento	LPS
	Periodismo e Información	
	Ciencias Sociales, periodismo e información sin mayores definiciones	
Ciencias naturales, matemáticas y estadística	Matemáticas y Estadísticas	GEEMP
	Medio Ambiente	LPS
	Ciencias Biológicas y Afines	
	Ciencias Físicas	GEEMP
	Ciencias Naturales, Matemáticas y Estadísticas sin mayor definición	
Educación	Educación	No STEM
Ingeniería, Industria y Construcción	Industria y Producción	GEEMP
	Arquitectura y Construcción	
	Ingeniería y Profesiones Afines	
Tecnología de la Información y la Comunicación (TIC)	Tecnología de la Información y la Comunicación (TIC)	GEEMP
Salud y Bienestar	Bienestar	LPS
	Salud	
Servicios	Servicios de Seguridad	No STEM
	Servicios de Higiene y Salud Ocupacional	

---

<b>Clasificación CINE - 2013</b>		<b>Categoría STEM</b>
<b>Área</b>	<b>Subárea</b>	
	Servicios personales	
	Servicios de Transportes	
Sin área definida	Sin área definida	

---

Fuente: Elaboración propia

## Anexos

### Anexo A. Identificación del estado del arte mediante una red de referencias

El presente anexo metodológico describe el procedimiento empleado para la identificación del estado del arte de un tópico de investigación a partir de la red de referencias que se construye al observar las citas y referencias de los artículos que a ella pertenecen.

Para dicha labor se emplea el proyecto OpenAlex<sup>25</sup>, una interfaz de código abierto que da acceso a un catálogo de artículos de investigación y publicaciones científicas junto a una descripción de dicho material, sus referencias y material que les cita.

Considerando la naturaleza referencial de la producción científica, es posible describirla en un gran grafo ( $\Psi$ ) en que cada artículo corresponde a un nodo y cada referencia que realiza describe un arco, de manera que cada área de investigación esté contenida en  $\Psi$  como un subgrafo de este.

Para identificar, mediante OpenAlex, el subgrafo ( $g$ ) se emplea una aproximación de bola de nieve que requiere, como entrada:

- Artículos raíz o semilla.
- Restricciones descriptivas de aquellos artículos que pertenecen a  $g$ .
- Un nivel máximo de búsqueda discreto e igual o mayor a 1.

Con ello, el algoritmo para la obtención de  $g$  se describe tal que:

1. Se adiciona cada artículo raíz al nivel 0.
2. Para cada artículo en el nivel actual:
  - a. Se obtienen los artículos que referencian a dicho artículo y cumplen las restricciones, para adicionarlos al nivel siguiente.
  - b. Se obtienen los artículos que dicho artículo referencia y cumplen las restricciones, para adicionarlos al nivel siguiente.

---

<sup>25</sup> Ver detalle en: <https://openalex.org/>

3. Si aún no se alcanza el nivel máximo, se vuelve al paso 2.

Con esto, se obtiene la información requerida para elaborar  $g$ , por lo que el procedimiento siguiente es identificar aquellos artículos que pertenezcan al estado del arte del tópico de investigación, es decir que sean relevantes producto de ser recientes y , a la vez, acumular conocimiento propio de dicho tópico.

Como se observa, dada la existencia de restricciones en la obtención de los datos sobre artículos y sus referencias, el subgrafo obtenido no describe ni pretende describir toda la producción científica relacionada al tópico de investigación, en cambio, se propone como una visión parcial y temporalmente delimitada de los avances en el área.

De este modo, para la evaluación de la relevancia y el conocimiento acumulado, es necesario considerar dicha propiedad de parcialidad del grafo, por lo que, en contraste a investigaciones previas que emplean métricas como el betweenness para evaluar la relevancia de un artículo, se propone, en cambio, emplear una aproximación de dos pasos para obtener, primero, la relevancia de cada artículo y luego distribuir hacia adelante dicha relevancia en forma de conocimiento acumulado.

Así, el primer paso consiste en computar la centralidad armónica<sup>26</sup> ( $H$ ) de cada artículo en un grafo con arcos orientados desde cada artículo hacia sus referencias presentes en  $g$ . Dado que lo anterior describe, primero, un grafo dirigido y, segundo, que dicha dirección está temporalmente delimitada (se asume que una referencia es siempre anterior al artículo que cita), lo esperable es que la relevancia (centralidad armónica) de un artículo correlacione positivamente con la antigüedad de este.

Luego, un segundo paso, consiste en replicar los principios que sustentan a la centralidad armónica como métrica de centralidad, para construir un indicador de conocimiento acumulado basado en un segundo grafo  $g'$  similar a  $g$ , pero con los arcos dirigidos en sentido opuesto (desde cada artículo hacia aquellos que lo referencian).

Así, este indicador de conocimiento acumulado se describe tal que:

---

<sup>26</sup> Revisar detalle en: <https://infoscience.epfl.ch/record/200525>

$$K(v) = \left( \sum_{u|u \neq v} H(u) \cdot \frac{1}{d(u, v)} \right) \cdot p$$

Donde  $v$  es el artículo (nodo) de interés en  $g'$ ,  $u$  es cada uno de los nodos de  $g'$ ,  $H$  es la función de centralidad armónica computada en el primer paso,  $d$  es una función de distancia entre dos nodos que describe la cantidad de nodos entre cada par (inicio inclusive) y  $p$  es un factor de penalización que permite disminuir el valor originalmente obtenido para situaciones como, por ejemplo, artículos con una gran cantidad de referencias, bajo el supuesto de que la ganancia marginal de información obtenida de una referencia inicial es decreciente, por ello se sugiere que  $p$  penalice inversamente respecto a, por ejemplo,  $\log(deg^+(v))$ , donde  $deg^+$  es el grado de salida de un nodo.

Tanto en el primer como en el segundo paso, el fundamento para emplear la centralidad armónica se desprende del supuesto de que la información presente en un artículo se distribuye en quienes el tópico de investigación a través de la evolución de este, disolviendo su aporte toda vez que dicha relación requiera de más pasos (artículos intermedios).

Así, se propone  $K$  como una función que permita ordenar descendentemente a los artículos de un tópico de investigación descritos en un grafo temporalmente parcial de dicho tópico, sin embargo, dado que un área de investigación está compuesta por subáreas de investigación que se diferencian entre sí por múltiples factores como la aproximación al fenómeno, sus métodos, disciplinas involucradas, entre otras, y que dichas subáreas son heterogéneas en su densidad, por lo que es necesario establecer una manera de captar dichas aproximaciones.

Para ello, se propone usar el algoritmo de detección de comunidades de *Louvain* sobre un tercer grafo de coocurrencia conceptual<sup>27</sup> para luego, para cada una de las comunidades captadas, considerar  $k$  artículos de mayor conocimiento acumulado como representativos de

---

<sup>27</sup> Donde la distancia entre dos artículos es equivalente a la distancia *hamming* para la lista de conceptos asociados a este.



dicha aproximación (donde  $k$  es un valor arbitrario por definirse en función de la cantidad de comunidades y de lo productiva que sea el tópico de investigación).

## **Anexo B. Estimación de nivel socioeconómico a nivel individual**

Para estimar el nivel socioeconómico de los estudiantes a nivel individual se dispone de dos aproximaciones a la vez incompletas.

Por un lado, los resultados de la medición de Vulnerabilidad Multidimensional del Estudiante (IVE) realizada por la Junta Nacional de Auxilio Escolar y Becas desde 2006 a través de la metodología IVE-SINAE y que da cuenta de aspectos psicosociales, culturales y ambientales que inciden en los niveles de desarrollo y bienestar de los estudiantes, considerando entre seis y diez dimensiones (en relación al ciclo de estudios) tales como su contexto socioeconómico familiar y comunal, su compromiso escolar, el contexto multiproblemático familiar, su historial académico, aspectos sobre la salud del estudiante o el apoyo y crianza que recibe<sup>28</sup>, generando así un índice estandarizado a nivel individual, escolar y comunal.

En segundo lugar, se dispone de la información de caracterización socioeconómica de los estudiantes que se inscriben al Sistema Único de Admisión, que informa entre otros detalles aspectos como el ingreso familiar y la cantidad de integrantes en la familia por cada estudiante.

Sin embargo, ninguna de las dos mediciones a nivel individual dispone de información para cada estudiante en la población estudiada, con menos de un 10% para el caso del IVE a nivel individual y de más o menos un 50% para los estudiantes inscritos a SUA. Además de esto, se dispone del IVE a nivel del establecimiento educacional<sup>29</sup> que, pese a disponer de información completa para el establecimiento de cada estudiante, no permite evaluar la variabilidad intra-escuela.

Por lo antes expuesto, se propone una estimación del nivel socioeconómico considerando, a la vez, el IVE de cada establecimiento y el ingreso per cápita de cada

---

<sup>28</sup> Más información en: <https://www.junaeb.cl/medicion-la-vulnerabilidad-ivm>

<sup>29</sup> Que fue calculado como la media geométrica del índice para cada ciclo.

estudiante, para luego realizar una imputación mediante aprendizaje automático para todo estudiante sin información disponible.

Para ello, considerando que el IVE es ascendente respecto a la vulnerabilidad del estudiante y el ingreso per cápita se distribuye de manera asimétrica con mayor concentración en tramos de ingreso inferiores, se estima un índice de situación socioeconómica individual (NSE) en que para cada estudiante que disponga de ambos datos informados, su NSE será calculado tal que para cada estudiante ( $i$ ):  $(1 - IVE_i) \cdot Ingreso_i$  donde  $IVE_i$  es el IVE del último establecimiento registrado en el historial de matrículas del estudiante e  $Ingreso_i$  es el ingreso per cápita calculado en función del ingreso bruto familiar y la cantidad de integrantes en el grupo familiar declarados en la primera inscripción al SUA realizada por el estudiante.

Luego, para imputar un NSE a los estudiantes sin información disponible, se realizó una imputación mediante un modelo de regresión basado en árboles con potenciación de gradientes basado en la edad de egreso, la modalidad de enseñanza, la comuna de residencia del estudiante, la dependencia desde la que egresó y si esta era o no rural.

Luego, para facilitar la interpretación de los resultados, se decidió aplicar una transformación de *Yeo-Johnson* para redistribuir los datos a una forma gaussiana y luego escalarla en el intervalo  $[0, 1]$ .

## Anexo C. Desempeño por modelo-hito

A continuación, se exponen detalles acerca de la elaboración y replicabilidad de los ocho modelos guía elaborados.

El primer modelo para identificar características relevantes para diferenciar entre ambos géneros en el egreso considera significativos todos los predictores ingresados salvo la comunidad de residencia del estudiante y si este egresó o no de educación artística, con una tasa de casos positivos (femeninos) de 822.526 (un 50,72% del total) y un desempeño a continuación descrito.

**Tabla 16. Desempeño por algoritmo para el primer modelo - hito**

Algoritmo	Recall		Precision		Puntaje de media geométrica
	Masculino	Femenino	Masculino	Femenino	
<i>LightGBM</i>	0,522505	0,633269	0,580099	0,577666	0,575218
<i>XGBoost</i>	0,524404	0,638083	0,584200	0,580480	0,578448
<i>CatBoost</i>	0,521574	0,637619	0,582565	0,578858	0,576684
<i>HGBClassifier</i>	0,522122	0,633522	0,580085	0,577567	0,575127

Fuente: Elaboración propia

Luego, el segundo modelo entrenado para clasificar a estudiantes que postulan vía SUA respecto a aquellos que sí lo hacen, con una cantidad de casos positivos de 1.043.387 (un 64,34% del total), que no considera significativos los mismos descriptores que el modelo antes comentado, con un desempeño que se describe a continuación.

**Tabla 17. Desempeño por algoritmo para el segundo modelo - hito**

Algoritmo	Recall		Precision		Puntaje de media geométrica
	No postula	Postula	No postula	Postula	
<i>LightGBM</i>	0,884802	0,682322	0,834212	0,766277	0,776994
<i>XGBoost</i>	0,885638	0,681519	0,833993	0,767370	0,776902
<i>CatBoost</i>	0,886708	0,679750	0,833393	0,768580	0,776362
<i>HGBClassifier</i>	0,884882	0,682003	0,834086	0,766318	0,776847

En tercer lugar, el modelo entrenado para distinguir, considerando a los estudiantes que postulan vía SUA, a aquellos que postulan al área GEEMP de aquellos que no descarta como significativos a los mismos descriptores antes mencionados, pero también al egreso desde modalidad PJA, la ruralidad del establecimiento y si este es o no de dependencia pública. Su desbalance es de 364.264 casos positivos que son un 62,98% del total y su desempeño se expone a continuación.

**Tabla 18. Desempeño por algoritmo para el segundo hito - submodelo GEEMP**

Algoritmo	Recall		Precision		Puntaje de media geométrica
	Postula a otras áreas	Postula a GEEMP	Postula a otras áreas	Postula a GEEMP	
<i>LightGBM</i>	0,845106	0,525107	0,751556	0,666040	0,666158
<i>XGBoost</i>	0,840753	0,530268	0,752631	0,662036	0,667699
<i>CatBoost</i>	0,844501	0,526489	0,751966	0,665756	0,666796
<i>HGBClassifier</i>	0,844776	0,525932	0,751808	0,665913	0,666552

Fuente: Elaboración propia

Luego, el modelo que busca diferenciar a estudiantes que se matriculan en la educación terciaria respecto a aquellos que no sólo descarta la relevancia de egresar desde enseñanza media artística y la comuna de residencia del estudiante, con un desbalance de 1.145.722 que es el 70,65% del total y un desempeño descrito a continuación.

**Tabla 19. Desempeño por algoritmo para el tercer modelo - hito**

Algoritmo	Recall		Precision		Puntaje de media geométrica
	No se matrícula	Se matrícula	No se matrícula	Se matrícula	
<i>LightGBM</i>	0,512142	0,866048	0,613735	0,810306	0,665986
<i>XGBoost</i>	0,514649	0,867633	0,617712	0,811378	0,668223
<i>CatBoost</i>	0,513326	0,867815	0,617425	0,810992	0,667436
<i>HGBClassifier</i>	0,512678	0,865525	0,613058	0,810382	0,666135

Fuente: Elaboración propia

En quinto lugar, considerando únicamente a los estudiantes que efectivamente se matriculan en la educación terciaria, se elabora un modelo que pretende diferenciar a aquellos que escogen carreras GEEMP respecto a aquellos que no, para ello resultan relevantes todos los descriptores ingresados salvo la comuna de residencia y si egreso o no desde educación artística, con un desbalance de 316.286 casos positivos que son el 27,61% del total y un desempeño que se expone a continuación.

**Tabla 20. Desempeño por algoritmo para el tercer hito - submodelo GEEMP**

Algoritmo	Recall		Precision		Puntaje de media geométrica
	Matricula en otras áreas	Matrícula en GEEMP	Matricula en otras áreas	Matrícula en GEEMP	
<i>LightGBM</i>	0,895264	0,452019	0,810562	0,622356	0,636136
<i>XGBoost</i>	0,893061	0,457343	0,811679	0,620214	0,639087
<i>CatBoost</i>	0,892635	0,459974	0,812348	0,620628	0,640770
<i>HGBClassifier</i>	0,895286	0,451477	0,810414	0,622128	0,635759

Fuente: Elaboración propia

El sexto modelo elaborado se construye para distinguir a estudiantes que se retiran desde la educación terciaria una vez matriculados. Su desbalance es de 435.540 casos positivos que son el 38,02% del total y su desempeño es tal como se expone en la tabla a continuación.

**Tabla 21. Desempeño por algoritmo para el cuarto modelo - hito**

Algoritmo	Recall		Precision		Puntaje de media geométrica
	Permanece	Se retira	Permanece	Se retira	
<i>LightGBM</i>	0,871320	0,376889	0,695100	0,642411	0,573052
<i>XGBoost</i>	0,868738	0,381100	0,695908	0,640409	0,575387
<i>CatBoost</i>	0,868140	0,384528	0,696936	0,641414	0,577775
<i>HGBClassifier</i>	0,871581	0,376157	0,694914	0,642436	0,572583

Fuente: Elaboración propia

El penúltimo modelo considera el hito del reingreso respecto a los estudiantes que se retiran de manera inicial o temprana y no registran titulación desde su primera carrera. Su

desbalance es de 269.846 casos positivos, que son el 66,23% del total y su desempeño se expone en la tabla a continuación.

**Tabla 22. Desempeño por algoritmo para el quinto modelo - hito**

Algoritmo	Recall		Precision		Puntaje de media geométrica
	No reingresa	Reingresa	No reingresa	Reingresa	
<i>LightGBM</i>	0,500892	0,839977	0,613935	0,768120	0,648642
<i>XGBoost</i>	0,504990	0,836820	0,611262	0,768919	0,650062
<i>CatBoost</i>	0,504454	0,838378	0,613253	0,769058	0,650324
<i>HGBClassifier</i>	0,500895	0,840047	0,614045	0,768138	0,648669

Fuente: Elaboración propia

Finalmente, para comprender la diferencia entre estudiantes que, habiendo reingresado, entran a una segunda carrera del área GEEMP, para ello el modelo entrenado incorpora a 66.341 casos positivos, ellos son el 24,58% del total y exhibe el siguiente desempeño.

**Tabla 23. Desempeño por algoritmo para el quinto hito - submodelo GEEMP**

Algoritmo	Recall		Precision		Puntaje de media geométrica
	Reingreso en otras áreas	Reingreso en GEEMP	Reingreso en otras áreas	Reingreso en GEEMP	
<i>LightGBM</i>	0,883889	0,573635	0,863726	0,617740	0,712056
<i>XGBoost</i>	0,886953	0,558989	0,860120	0,617940	0,704121
<i>CatBoost</i>	0,888566	0,553651	0,858887	0,619073	0,701386
<i>HGBClassifier</i>	0,884012	0,574014	0,863846	0,618142	0,712341

Fuente: Elaboración propia