

“La desigualdad socioeconómica y su impacto en el acceso a la educación superior”

Integrantes

- Joab Vivanco
- Javiera Bascur
- Alejandro Matheus
- Francisca Villablanca

Índice

1. Contexto y motivación	3
2. Objetivo	4
3. Datos	5
4. Preguntas de Investigación	<u>11</u>
5. Diseño Tentativo	11
6. Bibliografía	14

Contexto y motivación

Hoy en día, la educación en Chile se caracteriza por una profunda desigualdad, que afecta tanto el acceso como la calidad del proceso educativo entre los diferentes grupos socioeconómicos. Esta situación impide que la educación sea un derecho efectivo para todos los estudiantes, limitando las oportunidades de muchos.

La desigualdad educativa en el país proviene de una variedad de factores, como la calidad de la infraestructura, la disponibilidad de recursos educativos para los estudiantes, las condiciones socioeconómicas, la distribución de los docentes, el financiamiento público, la ubicación de los establecimientos, entre otros. Al comparar instituciones públicas y privadas, estas diferencias se vuelven aún más notables. A pesar de que ambas están bajo la supervisión del Ministerio de Educación, las medidas que este organismo implementa no logran abordar adecuadamente las desigualdades existentes. Como consecuencia, las diferencias se mantienen y afectan de manera desproporcionada a los sectores más vulnerables de la población.

En un contexto donde la educación es clave para el desarrollo social, la escasez de recursos y la desigualdad en las oportunidades educativas se convierten en barreras significativas. Las instituciones que reciben menos financiamiento enfrentan dificultades para proporcionar un entorno de aprendizaje adecuado, lo que limita la preparación de sus estudiantes. Esta situación crea un obstáculo considerable, ya que no todos los jóvenes tienen la misma capacidad para rendir la prueba de acceso a la educación superior de manera efectiva. Aunque todos los colegios del país (públicos, subvencionados y privados) utilizan diversas estrategias de preparación, las diferencias en recursos y apoyo educativo impactan directamente en las oportunidades de ingresar a una universidad.

“Cada año el 10% de los estudiantes abandona el sistema educativo, y de estos, el 80% lo hace debido a la pobreza. Además, el 68% de los egresados de la educación municipal logra acceder a la educación superior, en contraste con el 80% de los estudiantes de colegios subvencionados y el 95% de los de colegios privados” (Puroperiodismo, 2022).

Estas cifras evidencian la desigualdad que enfrenta el sistema educativo chileno, donde los estudiantes provenientes de familias con menos recursos tienen menos posibilidades de acceder a una educación superior de calidad.

Es por esta razón que nos motiva visibilizar este gran problema. Queremos destacar cómo las desigualdades socioeconómicas influyen en la desigualdad educativa, afectando el acceso a la educación superior y las diferentes oportunidades que tienen los estudiantes. Por ello, nuestro público objetivo incluye al gobierno, al Ministerio de Educación, a los establecimientos educativos y a otras entidades relevantes. Con el fin de generar conciencia sobre este tema, buscando impulsar cambios que permitan a todos los jóvenes acceder a una educación en condiciones más justas y equitativas.

Objetivo

El objetivo principal de este proyecto es analizar la relación entre el nivel socioeconómico, el tipo de institución educativa (pública, subvencionada o privada) y el desempeño académico de los estudiantes, explorando cómo la desigualdad social influye tanto el rendimiento académico como las oportunidades de acceso a la educación superior en condiciones de equidad.

Para ello, se llevará a cabo un estudio en el cual se analizará:

1. El desempeño académico en las diferentes pruebas de admisión a la educación superior entre estudiantes que asisten a distintos tipos de colegios (públicos, subvencionados y privados).
2. La variación del rendimiento académico en las pruebas de acceso a la educación superior en función del ingreso familiar per cápita de los estudiantes.
3. Las diferencias en las tasas de ingreso a la educación superior según el nivel socioeconómico, en comparación con las tasas de ingreso según las instituciones educativas a las que pertenecen los estudiantes.

Con el fin de responder a la pregunta, **¿Cuánto influye el nivel socioeconómico y el tipo de institución educativa en el acceso a la educación superior, y cómo se reflejan estas diferencias en el rendimiento académico de los estudiantes en la prueba de admisión?**

Permitiendo que el análisis pueda:

- Determinar si existe una relación significativa entre el nivel socioeconómico y el rendimiento académico en la prueba de admisión a la educación superior.
- Establecer cómo influye el tipo de institución educativa en el rendimiento académico de los estudiantes.
- Analizar qué factores socioeconómicos (como el ingreso familiar o la ubicación geográfica) tienen el mayor impacto en las oportunidades de acceso a la educación superior.

Con el fin de proporcionar datos e información resultantes de este estudio para que nuestra audiencia objetivo, como el gobierno, el Ministerio de Educación y los establecimientos educativos, pueda implementar diversas estrategias orientadas a desarrollar un plan que promueva la equidad en el acceso a la educación, mejorando el acceso a una educación de calidad. Esto se podría conseguir a través de iniciativas como el fortalecimiento de la formación docente, la optimización de la distribución de recursos y la creación de programas de apoyo académico dirigidos a estudiantes de sectores más vulnerables, asegurando así un futuro más equitativo para todos los jóvenes chilenos.

Datos (<https://github.com/javibascur/Proyecto-Cs-Datos>)

Este proyecto abarca el análisis de datos de postulantes a la educación superior en el período comprendido entre 2019 y 2023. En el cual, se utilizarán cuatro tipos principales de datasets (csv) provenientes de una fuente publica, los cuales son:

1. **Inscripción:** Datos relacionados con el registro de los postulantes a la Prueba de Acceso a la Educación Superior.

2. **Rendición:** Datos sobre los resultados de las pruebas y evaluaciones rendidas por los postulantes.
3. **Postulación:** Datos sobre las preferencias de carreras y el estado de las postulaciones.
4. **Matrícula:** Datos sobre las matrículas efectivas en universidades, incluyendo los tipos de vía y puntajes ponderados.

Cada año cuenta con un conjunto de archivos para estos cuatro tipos de datos, los cuales se extraerán y convertirán a dataframes para su posterior análisis y así responder a la pregunta planteada en el objetivo de nuestro proyecto. Estos archivos están disponibles en el repositorio de GitHub del proyecto, junto con los respectivos dataframes.

1. Datos obtenidos de la inscripción a la Prueba de Acceso a la educación Superior (portal de datos abiertos - DEMRE)

Este conjunto de datos contiene información detallada sobre las inscripciones de los estudiantes a la Prueba de Acceso a la Educación Superior. Las variables abarcan aspectos personales, académicos, socioeconómicos y geográficos de los postulantes, lo que permite obtener un perfil completo de cada uno de ellos.

Las columnas que se utilizarán para crear un dataframe son:

- **ID_AUX (str)** Identificador único del postulante.
- **ANYO_PROCESO (int):** Año correspondiente al proceso de admisión.
- **SEXO (int):** Sexo del postulante,
- **GRUPO_DEPENDENCIA (float):** Tipo de financiamiento del establecimiento de egreso.
- **FECHA_NACIMIENTO (int):** Mes y año de nacimiento del postulante.
- **ANYO_EGRESO (float):** Año en que el postulante egresó de la educación media.
- **CODIGO_REGION_D (int):** Código de la región de domicilio del postulante.
- **CODIGO_COMUNA_D (int):** Código de la comuna de domicilio del postulante.

- **INGRESO_PERCAPITA_GRUPO_FA (int):** Ingreso per cápita del grupo familiar del postulante.

	ID_aux	ANYO_PROCESO	SEXO	GRUPO_DEPENDENCIA	FECHA_NACIMIENTO	ANYO_EGRESO	CODIGO_REGION_D	CODIGO_COMUNA_D	INGRESO_PERCAPITA_GRUPO_FA
0	id_9265592376510	2023	1	4.0	92004	2022.0	15	15101	2
1	id_9492594679209	2023	1	4.0	62004	2021.0	15	15101	2
2	id_9325593372588	2023	1	4.0	22005	2022.0	15	15101	2
3	id_3443934474359	2023	1	4.0	61992	2010.0	15	15101	1
4	id_9355593175583	2023	1	4.0	122004	2022.0	15	15101	99
...
293860	id_9162591776233	2023	1	NaN	112003	2022.0	13	13125	99
293861	id_6620966272093	2023	2	NaN	61998	2016.0	4	4101	1
293862	id_8886588078640	2023	2	NaN	61999	2019.0	9	9101	2
293863	id_3253532675336	2023	2	NaN	102002	2021.0	1	1101	8
293864	id_8843588774312	2023	1	NaN	101998	2019.0	8	8105	99

293865 rows × 9 columns

Figura 1. Dataframe del dataset inscripción (2023)

2. Datos obtenidos de la rendición a la Prueba de Acceso a la educación Superior (portal de datos abiertos - DEMRE)

Este conjunto de datos contiene información detallada sobre el perfil académico de cada estudiante y los resultados del desempeño en cada una de las pruebas que rindió el alumnado. Las variables se dividen principalmente en cuatro categorías, las cuales son la información académica, los resultados de cada prueba, la información geográfica y académica del estudiante.

Las columnas que se utilizarán para crear un dataframe son:

- **ID_AUX (str):** Identificador único del postulante.
- **PROMEDIO_NOTAS (str):** Promedio de notas de la educación media del postulante.
- **PTJE_NEM (int):** Puntaje estándar asignado al promedio de notas (NEM) del postulante.
- **PTJE_RANKING (int):** Puntaje estándar asignado al ranking de notas del postulante.
- **CLEC_REG_ACTUAL (float):** Puntaje en la prueba de Competencia Lectora actual del postulante.

- **MATE1_REG_ACTUAL (float):** Puntaje en la prueba de Competencia Matemática 1 actual del postulante.
- **MATE2_REG_ACTUAL (float):** Puntaje en la prueba de Competencia Matemática 2 actual del postulante.
- **HCSOC_REG_ACTUAL (float):** Puntaje en la prueba de Historia actual del postulante.
- **CIEN_REG_ACTUAL (float):** Puntaje en la prueba de Ciencias actual del postulante.
- **GRUPO_DEPENDENCIA (float):** Tipo de financiamiento del establecimiento de egreso.
- **CODIGO_REGION (float):** Código de la región del establecimiento de egreso del postulante.
- **CODIGO_COMUNA (float):** Código de la comuna del establecimiento de egreso del postulante.

	ID_aux	GRUPO_DEPENDENCIA	CODIGO_REGION	CODIGO_COMUNA	PROMEDIO_NOTAS	PTJE_NEM	PTJE_RANKING	CLEC_REG_ACTUAL	MATE1_REG_ACTUAL	MATE2_REG_ACTUAL	HCSOC_REG_ACTUAL	CIEN_REG_ACTUAL
0	id_0000900170003	3.0	13.0	13401.0	5,03	429	429	0.0	402.0	351.0	423.0	0.0
1	id_0000900470024	3.0	13.0	13107.0	5,9	689	766	475.0	0.0	0.0	0.0	0.0
2	id_0000900670068	2.0	1.0	1101.0	5	415	415	536.0	413.0	0.0	489.0	0.0
3	id_0001900070103	2.0	14.0	14202.0	5,6	599	599	589.0	437.0	0.0	0.0	478.0
4	id_0001900070149	3.0	14.0	14101.0	5,28	506	517	0.0	0.0	0.0	0.0	0.0
...
293860	id_9999599979968	2.0	14.0	14106.0	5,98	708	785	NaN	NaN	NaN	NaN	NaN
293861	id_999990948930	2.0	13.0	13129.0	6,03	724	754	0.0	0.0	0.0	0.0	0.0
293862	id_999994942103	3.0	4.0	4301.0	5,73	637	652	421.0	467.0	472.0	400.0	481.0
293863	id_999997943080	NaN	NaN	NaN	0	0	0	775.0	437.0	0.0	521.0	582.0
293864	id_9999999879976	3.0	13.0	13119.0	5,85	675	715	0.0	0.0	0.0	0.0	0.0

293865 rows × 12 columns

Figura 2. Dataframe del dataset rendición (2023)

3. Datos obtenidos de las postulaciones a Instituciones de Educación Superior (portal de datos abiertos - DEMRE)

Este conjunto de datos ofrece un registro exhaustivo de las postulaciones de los estudiantes, abarcando programas especiales como BEA y PACE, así como los puntajes obtenidos y el estado de las postulaciones.

Las columnas que se utilizarán para crear un dataframe son:

- **ID_AUX (str):** Identificador único del postulante.
- **SITUACION_POSTULANTE (str):** Código único que indica la situación del postulante respecto a sus postulaciones en el programa BEA
- **PTJE_PREFXX (str):** Puntaje ponderado del postulante.
- **SIT_POSTULANTE_BEA (str):** Código que indica la situación del postulante respecto a sus postulaciones en el programa BEA
- **PTJE_PREFXX_BEA (float):** Puntaje ponderado de estudiantes con postulación al programa BEA.
- **SIT_POSTULANTE_PACE (str):** Código que indica la situación del postulante respecto a sus postulaciones en el programa PACE.
- **PTJE_PREFXX_PACE (float):** Puntaje ponderado de estudiantes con postulación al programa PACE.

	ID_aux	SITUACION_POSTULANTE	PTJE_PREF_01	SITUACION_POSTULANTE_BEA	PTJE_PREF_01_BEA	SITUACION_POSTULANTE_PACE	PTJE_PREF_01_PACE
0	id_0001900070103	P	543,9	NaN	NaN	NaN	NaN
1	id_0001900170136	P	794,95	NaN	NaN	NaN	NaN
2	id_0001900670175	P	570,1	NaN	NaN	NaN	NaN
3	id_0001900870116	P	719,9	NaN	NaN	NaN	NaN
4	id_0002900370264	P	522,9	NaN	NaN	NaN	NaN
...
86059	id_9999599979931	P	506,9	NaN	NaN	NaN	NaN
86060	id_9999599979949	P	539,3	NaN	NaN	NaN	NaN
86061	id_9999599979953	P	684,9	NaN	NaN	NaN	NaN
86062	id_9999599979968	P	593,3	NaN	NaN	NaN	NaN
86063	id_999997943080	P	533,6	NaN	NaN	NaN	NaN

6064 rows × 7 columns

Figura 3. Dataframe del dataset Postulaciones (2023)

4. Datos obtenidos de las matrículas a Instituciones de Educación Superior (portal de datos abiertos – DEMRE)

Este conjunto de datos contiene información sobre los postulantes que se han matriculado en universidades a través de diferentes vías y tipos de matrícula. Los datos se centran en las universidades, las carreras y las características de la matrícula de los estudiantes, además del puntaje ponderado.

Las columnas que se utilizarán para crear un dataframe son:

- **ID_AUX (str):** Identificador único del postulante.
- **VIA (int):** Vía de admisión de la matrícula del postulante
- **PTJE_PONDERADO (float):** Puntaje ponderado con el que el postulante se matriculó.

	ID_aux	VIA	PTJE_POND
0	id_267832678788	1	513.60
1	id_016130170616	1	565.85
2	id_514305173430	1	724.70
3	id_042500470254	1	697.60
4	id_699806974987	1	722.00
...
111292	id_8198841548154	1	591.45
111293	id_7666786944704	1	580.50
111294	id_7666786243268	1	733.50
111295	id_7666786649432	1	593.40
111296	id_7666786640438	1	677.30

111297 rows × 3 columns

Figura 4. Dataframe del dataset matrículas (2023)

Preguntas de investigación

- ¿Cuál es el tipo de colegio con la mayor tasa de abandono en el proceso de postulación universitaria?
- ¿Qué tipo de institución educativa (privada, subvencionada o pública) presenta la mayor tasa de acceso a la educación superior entre sus egresados?
- ¿Qué tan alta es la tasa de apoyo de la PACE en el acceso a la educación superior para los jóvenes de sectores con alta vulnerabilidad económica?
- ¿Cuáles son los resultados promedio en la prueba de acceso a la educación superior según el tipo de colegio en cada comuna?
- ¿En qué comunas se concentran los diferentes tipos de colegios y cómo se relaciona esto con los resultados de egreso de los estudiantes?

Diseño Tentativo

Con el propósito de evaluar y analizar el impacto de la desigualdad socioeconómica en el rendimiento académico de los estudiantes y su acceso a la educación superior, se aplicarán diferentes métodos computacionales y estadísticos, los cuales serán detallados en las secciones que siguen.

1. Limpieza de datos

Los datasets de inscripción, rendición de pruebas, postulaciones y matrículas serán transformados y limpiados, manejando posibles errores, valores nulos, datos atípicos y cualquier incoherencia encontrada. Se emplearán estadísticas descriptivas para

identificar los valores faltantes o nulos y se determinará el impacto de estos en el análisis. Posteriormente, se utilizarán técnicas de normalización e imputación según corresponda.

Para que el algoritmo funcione de manera correcta, se utilizará Pandas para detectar e identificar duplicados, eliminándolos cuando sea necesario. También se verificarán los datos categóricos para asegurar que las categorías estén correctamente definidas y codificadas. Esto implicará la revisión de variables como el tipo de colegio, género y nivel socioeconómico, corroborando que las categorías se mantengan consistentes en todos los registros.

En caso de que los datos estén desequilibrados (por ejemplo, si hay una cantidad desproporcionada de estudiantes de un tipo de colegio frente a otro) se considerará la aplicación de técnicas como el sobremuestreo o submuestreo, para balancear los grupos y para que los modelos predictivos puedan generar resultados confiables. De esta manera se obtendrán los datasets filtrados con la información necesaria para su posterior análisis.

2. Análisis

Para comprender cómo la desigualdad socioeconómica en los distintos establecimientos educacionales puede afectar el rendimiento académico, se utilizarán varios enfoques de análisis para así, poder identificar patrones y relaciones significativas entre las variables.

En primer lugar, se aplicarán técnicas de estadística descriptiva para generar un perfil general de los estudiantes. Se analizarán factores como el tipo de institución (público, privado, subvencionado), el ingreso per cápita familiar y los puntajes obtenidos en pruebas estandarizadas. Con Pandas y NumPy se calcularán estadísticas como la media, mediana y desviación estándar para observar la dispersión de los datos.

Para predecir cómo los factores socioeconómicos influyen en el acceso a la educación superior, se utilizarán modelos de regresión logística, lo que ayudará a estimar la probabilidad de que un estudiante acceda a la universidad, considerando variables como el ingreso familiar, el tipo de colegio y los puntajes en pruebas.

Para evaluar la relación entre las variables cuantitativas, se generarán matrices de correlación y se aplicarán pruebas estadísticas, como el coeficiente de correlación de

Pearson. Esto permitirá determinar cómo los factores, tales como los ingresos per cápita familiar y los puntajes obtenidos en las pruebas influyen en el acceso a la educación superior.

Con ANOVA se realizarán comparaciones entre distintos grupos de estudiantes, lo que permitirá identificar si existen diferencias significativas en el rendimiento académico entre alumnos de diferentes tipos de colegios y comunas, como también en el acceso a la universidad.

3. Visualización de Datos

Para facilitar la comprensión de los datos, se emplearán librerías en Python, tales como Matplotlib (probablemente junto con complementos como “cycler” o “seaborn”), Bokeh, Pandas y NumPy. El uso de Pandas nos permitirá organizar los datos principales de estudio y manejo, facilitando la manipulación de estos. Además, con la ayuda de NumPy, realizaremos cálculos numéricos complejos, lo que permitirá manipular los datos de manera eficiente. Las otras librerías se centran en graficar estos datos de diversas formas, proporcionando variedad y un mejor entendimiento del tema. Por ejemplo, la librería Bokeh permite crear gráficos interactivos, ofreciendo al público una experiencia personal con la información que planeamos proporcionar. Matplotlib y Seaborn facilitarán la creación de gráficos estadísticos, como gráficos de dispersión, histogramas y gráficos de caja, que ayudarán a visualizar la distribución y las relaciones entre las variables. La librería Bokeh, por su parte, permitirá crear gráficos interactivos e

Bibliografía

- Villalobos, A. (2022, noviembre 2). Desigualdad educativa: El problema de base que no se soluciona con la PAES. Puroperiodismo.
<https://www.puroperiodismo.cl/desigualdad-educativa-el-problema-de-base-que-no-se-soluciona-con-la-paes/>
- Good Neighbors Chile. (2023). ¿Qué es la desigualdad educativa?. Good Neighbors Chile. <https://goodneighbors.cl/desigualdad-educativa/>
- Departamento de Evaluación, Medición y Registro Educacional (DEMRE). (n.d.). Portal bases de datos. DEMRE. <https://demre.cl/portales/portal-bases-datos>
- Bokeh. (2024, septiembre 26). Bokeh: Interactive visualization. PyPi.
<https://pypi.org/project/bokeh/>
- Ciencia de Datos. (n.d.). ANOVA: Prueba de hipótesis. Ciencia de Datos.
https://cienciadedatos.net/documentos/19_anova
- IBM. (n.d.). One-sample T-test. IBM Docs. <https://www.ibm.com/docs/es/spss-statistics/saas?topic=tests-one-sample-t-test>
- Pure Storage. (2023). What is data preprocessing?. Pure Storage.
<https://www.purestorage.com/es/knowledge/what-is-data-preprocessing.htm>