



**Universidad de Concepción
Campus Los Ángeles
Escuela de Educación**

**Construcción y evaluación de un modelo de predicción del
rendimiento académico en Física y Química a partir de
puntajes de ingreso y pruebas diagnósticas en estudiantes
de Educación General Básica de la Universidad de
Concepción, Campus Los Ángeles.**

**Tesis presentada a la escuela de educación de la Universidad de
Concepción para optar al título profesional de Profesor de Educación
General Básica con mención en Matemática y Ciencias Naturales.**

Por: Axcel Yermain Almuna Basso

Profesor guía: Dr. Francisco Antonio

Brovelli Sepúlveda

Comisión evaluadora:

Mg. Álvaro Moya Oliva

Mg. Fabián Cifuentes Rebolledo

**Los Ángeles, Chile
2025**

© 2025, Axcel Yerman Almuna Basso.

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo la cita bibliográfica del documento.

Agradecimientos.

En primera instancia, deseo agradecer a Dios por permitirme alcanzar esta meta y por brindarme la oportunidad de coincidir con personas valiosas a lo largo de mi formación. Mi gratitud se extiende a mis profesores y compañeros de diversos cursos, carreras y universidades; cada uno de ellos enriqueció mi trayectoria, aportando significativamente a mi crecimiento tanto profesional como personal.

Expreso mi más profundo agradecimiento a mis padres, Axcel y Carolina, y a mi hermano, Máximo, por brindarme siempre el apoyo incondicional y el impulso necesario para perseguir mis sueños. Cada uno de sus consejos ha sido escuchado y atesorado, viendo hoy los frutos de su constante guía y amor.

A mi amiga, compañera y pareja, Camila, quien ha sido un pilar fundamental en este tránsito universitario. Gracias por ser esa luz que transformó mis días grises y por confiar en mis capacidades, incluso en aquellos momentos en los que yo mismo dudaba de ellas. Tu presencia ha marcado toda la diferencia en este camino.

Asimismo, agradezco a mis amigos Álvaro y Gustavo por su lealtad inquebrantable. Gracias por estar presentes en las malas y en las buenas, demostrando el valor de la verdadera amistad cuando más se necesita.

Finalmente, quiero manifestar mi gratitud hacia mi profesor guía. Su confianza inicial en mi potencial académico fue el motor para desarrollar este trabajo de título con éxito. A pesar de los desafíos y altibajos, esta investigación pudo concretarse gracias a su orientación y compromiso.

Índice

Resumen	5
Abstract	5
Introducción	6
Capítulo I: Planteamiento del problema y justificación de la Investigación.	
1.1. Planteamiento del problema	7
1.1.1. Antecedentes generales	8
1.2. Pregunta de investigación	9
1.3. Objetivos de la investigación	9
1.3.1. Objetivo general	9
1.3.2. Objetivos específicos	10
1.4. Justificación del problema	10
Capítulo II: Marco Teórico.	
2.1. Factores académicos que inciden en el rendimiento en la educación superior	12
2.2. Factores académicos que inciden en el rendimiento académico en las asignaturas de Física y Química	14
2.3 Modelos predictivos del rendimiento académico	15
Capítulo III: Marco Metodológico.	
3.1. Tipo o Enfoque de la Investigación	18
3.2. Diseño de Investigación	18
3.3. Dimensión Temporal	19
3.4. Alcance de la Investigación	19
3.5. Población	19
3.6. Participantes del Estudio	19
3.7. Unidad de Análisis	20
3.8. Recolección de Datos	20
3.9. Técnica de Análisis de Datos	21
3.10. Consideraciones Éticas	22
Capítulo IV: Resultados.	
4.1 Análisis estadístico preliminar	23
4.2 Pruebas de normalidad	34
4.3 Análisis no paramétrico de las notas de Física y Química según contexto escolar	34
4.4 Modelado predictivo	36
4.4.1 Análisis de correlación no paramétricas de variables predictoras	36
4.4.2 Análisis de colinealidad de variables predictoras	38
4.4.3 Entrenamiento de algoritmos predictivos	39
4.4.4 Resultados entrenamiento para FIS y QUI	39
Capítulo V: Discusión.	43
Capítulo VI: Conclusiones.	45
Limitaciones y proyecciones del estudio.	47
Referencias.	48

Resumen.

Esta investigación construyó y evaluó un modelo predictivo del rendimiento en Física y Química para estudiantes de Educación General Básica (cohortes 2018-2024). Mediante un enfoque cuantitativo, se emplearon como predictores los antecedentes de ingreso y resultados diagnósticos institucionales. Los hallazgos demuestran que las habilidades matemáticas y los puntajes en ciencias tienen mayor incidencia en el desempeño, revelando además un dominio previo insuficiente. Se determinó que la capacidad de predicción es diferenciada: Química presenta una relación consistente que se fortalece con el éxito en Física, mientras que en esta última la relación es más débil. En conclusión, el modelo se valida como una herramienta de alerta temprana para identificar estudiantes en riesgo, siendo una ayuda para el diseño de estrategias de nivelación que garanticen el éxito en la formación inicial docente.

Palabras clave: Rendimiento académico, modelo predictivo, factores académicos de ingreso, Educación General Básica, Física, Química, alerta temprana, pruebas diagnósticas, formación inicial docente.

Abstract.

This research constructed and evaluated a predictive model of performance in Physics and Chemistry for students in Basic General Education (cohorts 2018-2024). Using a quantitative approach, the model employed entry history and institutional diagnostic results as predictors. The findings demonstrate that mathematical skills and science scores have a greater impact on performance, also revealing insufficient prior mastery. The predictive capacity was found to be differentiated: Chemistry shows a consistent relationship that is strengthened by success in Physics, while the relationship with Physics is weaker. In conclusion, the model is validated as an effective early warning tool for identifying at-risk students, facilitating leveling strategies that ensure success in initial teacher training.

Keywords: Academic performance, predictive model, academic factors of admission, Basic General Education, Physics, Chemistry, early warning, diagnostic tests, initial teacher training.

Introducción.

En el escenario actual de la educación superior, el rendimiento académico en las asignaturas de Ciencias Naturales representa un desafío crítico y una preocupación constante para las instituciones de formación docente. Particularmente en la Universidad de Concepción, se ha identificado que un número significativo de estudiantes de la carrera de Educación General Básica enfrenta dificultades considerables en las asignaturas de Física y Química durante su segundo año de formación. Estas disciplinas no solo demandan un alto nivel de razonamiento lógico y comprensión abstracta, sino que también requieren un dominio de habilidades matemáticas que suele ser superior al exigido en otras áreas como la Biología. La presente investigación nace de la necesidad de abordar esta problemática mediante un enfoque preventivo, ya que el problema central radica en que, a pesar de la alta exigencia cognitiva de estas asignaturas, la institución no cuenta actualmente con una herramienta que permita identificar de manera anticipada a los estudiantes que presentan un mayor riesgo de bajo rendimiento. Esta falta de detección temprana limita la capacidad de implementar estrategias de acompañamiento y nivelación oportunas que aseguren el éxito académico y reduzcan el riesgo de deserción. El estudio se fundamenta en la premisa de que los factores académicos de ingreso, como los puntajes en las pruebas de selección universitaria y los resultados de las evaluaciones diagnósticas institucionales, actúan como indicadores del nivel de dominio conceptual y las habilidades cognitivas con las que el estudiante inicia su trayectoria universitaria. Bajo esta perspectiva, se plantea como objetivo general construir y evaluar un modelo de predicción del rendimiento académico en Física y Química para los estudiantes de Educación General Básica del Campus Los Ángeles. Para alcanzar este propósito, la investigación se estructuró en varias etapas: primero, la caracterización académica de los estudiantes de las cohortes 2018-2024; segundo, la identificación de las variables de ingreso con mayor capacidad predictiva; y finalmente, el desarrollo y validación de un modelo estadístico capaz de estimar el rendimiento académico futuro. A través de un enfoque cuantitativo y un diseño no experimental correlacional, este trabajo busca no solo aportar evidencia empírica sobre la relación entre los antecedentes escolares y el éxito universitario, sino también entregar a la Universidad de Concepción una herramienta funcional de alerta temprana que contribuya a fortalecer el perfil de egreso de los futuros docentes, garantizando que cuenten con los conocimientos disciplinares sólidos necesarios para impartir una educación de calidad en el área de las Ciencias Naturales.

Capítulo I: Planteamiento del problema y justificación de la Investigación.

1.1 Planteamiento del problema.

En la educación superior, el bajo rendimiento académico en las asignaturas de las Ciencias Naturales constituye una preocupación permanente, particularmente en carreras relacionadas con la formación docente. En la Universidad de Concepción, se ha observado que existe un número no menor de estudiantes de la carrera de Educación General Básica que enfrentan dificultades en las asignaturas de Física y Química, asignaturas que se imparten durante el segundo año de formación. Estas asignaturas representan un eje fundamental en la preparación del futuro docente, al requerir habilidades de razonamiento lógico, comprensión abstracta y aplicación de conceptos científicos básicos, características que las convierten en áreas de alta demanda cognitiva y conceptual. La literatura en didáctica de las ciencias ha señalado que Física y Química suelen presentar mayores niveles de dificultad para los estudiantes, debido al alto grado de abstracción y formalización que exigen, lo que impacta directamente en el aprendizaje y el rendimiento académico (Quintanal Pérez, 2023).

La investigación parte del supuesto de que los factores académicos de ingreso, como los resultados de las pruebas de admisión universitaria y las pruebas diagnósticas institucionales, podrían actuar como predictores del rendimiento académico en dichas asignaturas. Estos parámetros reflejan el nivel de dominio de los contenidos escolares y las habilidades cognitivas con las que el estudiante inicia su trayectoria universitaria, constituyendo un referente inicial relevante, para anticipar el desempeño en asignaturas de alta demanda cognitiva, como Física y Química, aun cuando estas mediciones no capturan la totalidad de los factores que inciden en el rendimiento académico, dado que el rendimiento académico universitario se ve influido por variables de presagio asociadas al rendimiento previo y a las competencias académicas iniciales del estudiante (López Aguado y López Alonso, 2013). Al centrarse exclusivamente en estas asignaturas, se enfatiza la relevancia del salto conceptual y matemático que estas exigen, el cual resulta mayor en comparación con otras asignaturas del área de las Ciencias Naturales, como la asignatura de Biología, y que, por ende, tienden a presentar mayores dificultades para los estudiantes.

En relación con estos hechos, se ha indagado al respecto, llegando a la conclusión que la Universidad de Concepción, no cuenta con una herramienta capaz de identificar y anticiparse a estos fenómenos. Por ello, surge la necesidad de generar un modelo predictivo, con la finalidad de que éste, en

un futuro, pueda ser utilizado por la institución para generar estrategias de apoyo, dado que, muchos estudiantes avanzan curricularmente presentando dificultades que, posiblemente, podrían mitigarse si se genera el acompañamiento correspondiente en el rendimiento académico. Esta investigación no busca señalar ni analizar causas profundas del bajo rendimiento, sino construir un puente entre los datos académicos de ingreso y el resultado concreto, mediante un modelo predictivo construido con información objetiva, para identificar los alumnos susceptibles a un bajo rendimiento académico, reconociendo que las variables consideradas son acotadas y que el rendimiento académico corresponde a un fenómeno complejo y multifactorial.

1.1.1. Antecedentes generales.

El rendimiento académico en educación superior, según Lerner (2010), es un acontecimiento que puede ser analizado desde tres dimensiones: como un producto cuantitativo (calificaciones), un juicio de evaluación cualitativo (evaluación mixta), y un proceso continuo que refleja la interacción entre la enseñanza y el aprendizaje. Esta visión es complementada por Edel (citado por Ortiz Ojeda y Canto Herrera, 2013), quien define el rendimiento como la evaluación del conocimiento, pero ampliando la mirada, relacionando las notas obtenidas en las asignaturas con factores contextuales, los cuales presentan una relación de incidencia directa (García Arista *et al.*, 2024; Vera Sagredo *et al.*, 2021). En este sentido, mientras Lerner se enfoca en la naturaleza misma del rendimiento, Edel y Martínez Otero Pérez, V., (2021), profundizan en los elementos del entorno que lo moldean, demostrando que su comprensión requiere un análisis articulado entre el rendimiento cuantificable (calificaciones) y el establecimiento de formación escolar del estudiante (Campos Palacios, 2025).

Analizar este fenómeno, resulta esencial para identificar los posibles desafíos de la educación superior, inherentes al rendimiento académico en educación superior. Lara *et al.* (2014) afirman que las calificaciones obtenidas por los estudiantes en los primeros años de educación superior son el reflejo de una serie de dificultades académicas que, por diferentes motivos, influyen significativamente en la nota final de una asignatura. De la misma forma, Duche, Gallardo y González (2023), señalan que, la transición hacia la educación superior es un desafío que repercute en el rendimiento de los estudiantes, por la simple razón que éste implica una demanda mayor en la exigencia académica, y es aquí, donde se suelen evidenciar diferencias en la preparación previa del estudiante, afectando directamente el rendimiento académico durante los primeros años de educación superior. Por su parte, Soza Mora (2021) concluye que

el rendimiento académico es el resultado de una serie de competencias cognitivas previas del estudiante, las cuales se traducen, posteriormente, en resultados cuantitativos (calificaciones).

De acuerdo con estos antecedentes, es importante considerar el historial académico de los estudiantes, previo al ingreso a la educación superior. Algo similar fue lo planteado por Contreras (2021), quien, además, considera la relevancia del puntaje obtenido en la prueba de matemáticas, en el proceso de admisión universitaria, y la incidencia de éste en el rendimiento académico en las asignaturas que requieren un pensamiento lógico matemático.

A partir de estas consideraciones, se deduce la relevancia de focalizar la atención en los primeros años de la formación docente (Escuela de Profesores del Perú, 2024) y además la importancia que esta etapa debería tener para la Universidad de Concepción, puesto que, para la institución es fundamental el desarrollo disciplinar en los futuros docentes, ya que de acuerdo a su perfil de egreso, (Universidad de Concepción, 2024) buscan garantizar una educación de calidad, es decir, que los docentes egresados, demuestren competencias profesionales que integren conocimientos disciplinares sólidos de las Ciencias Naturales (Bórquez Mella *et al.*, 2023).

Es por ello, la pertinencia de focalizar el análisis en los factores académicos de ingreso que, pueden predecir el rendimiento académico de los estudiantes en asignaturas que requieren un alto nivel cognitivo en Educación General Básica, como Física y Química (Bastías Bastías e Iturra Herrera, 2022).

1.2 Pregunta de investigación

¿En qué medida los factores académicos de ingreso permiten predecir el rendimiento académico de los estudiantes que cursan las asignaturas de Física y Química en la carrera de Educación General Básica de la Universidad de Concepción, Campus Los Ángeles?

1.3.1. Objetivo General

Evaluar un modelo de predicción del rendimiento académico en las asignaturas de Física y Química, considerando como variables predictivas, factores académicos de ingreso de estudiantes de la carrera de Educación General Básica de la Universidad de Concepción Campus Los Ángeles.

1.3.2. Objetivos Específicos

1. Caracterizar académicamente a los estudiantes que ingresan a primer año de la carrera de Educación General Básica.
2. Identificar los factores académicos de ingreso que permitan construir un modelo de predicción de rendimiento académico en las asignaturas de física y química.
3. Evaluar la relación entre los factores académicos de ingreso identificados y el rendimiento académico en las asignaturas de Física y Química.
4. Desarrollar y evaluar un modelo estadístico predictivo que estime el rendimiento académico en las asignaturas de Física y Química.

1.4. Justificación del problema

Lograr un buen rendimiento académico en una carrera universitaria constituye un desafío permanente para los estudiantes, razón por la cual las instituciones de educación superior realizan esfuerzos sistemáticos orientados a fortalecer el desarrollo de habilidades cognitivas de orden superior, tales como el razonamiento lógico, el pensamiento abstracto y la capacidad de aplicar conocimientos en contextos diversos. No obstante, este desafío no se distribuye de manera homogénea entre las distintas asignaturas del currículo, ya que existen áreas disciplinares que, por su naturaleza conceptual y formal, presentan mayores niveles de complejidad y exigencia académica. En este contexto, asignaturas como Física y Química en la carrera de Educación General Básica representan desafíos significativamente mayores para los estudiantes, debido a la alta carga de abstracción, formalización matemática y articulación de múltiples conceptos que demandan. La evidencia especializada ha señalado que estas disciplinas suelen constituirse en asignaturas críticas dentro de los planes de estudio, afectando el desempeño académico y el progreso curricular de los estudiantes, lo que resulta especialmente relevante en la formación inicial docente, donde el dominio de los fundamentos científicos es clave para una enseñanza de calidad en los niveles escolares (De Luca *et al.* 2025).

Estas asignaturas en particular requieren de un alto nivel de pensamiento matemático, a diferencia de la asignatura de Biología. Si bien, existe una relación entre las tres ciencias (Física, Química y Biología), esta última, requiere competencias más conceptuales, la cual es una competencia propia del lenguaje (Salgado Díaz, 2022). Con esto, cursar dichas asignaturas (Física y Química), conllevan una dificultad y exigencia mayor para los futuros docentes (Landin García *et al.*, 2024), lo que se ve

complejizado por la deficiente formación formal de contenidos a lo largo de todo el ciclo escolar (Graciela Valle *et al.*, 2014). Estas deficiencias de formación disciplinaria en la educación escolar se manifiestan a través de los puntajes obtenidos en las pruebas del proceso de admisión universitaria y en las evaluaciones diagnósticas de ingreso a la carrera, siendo un valor importante de considerar en la construcción de un modelo predictivo del rendimiento académico en la educación superior (González Sanzana y Arce Secul, 2021).

Ahora bien, un estudiante puede presentar dificultades en alguna de las asignaturas, tanto en Física como en Química, sin embargo, su bajo rendimiento no siempre se debe a factores inherentes a él, sino que puede estar influenciado por factores extrínsecos, como la alta exigencia y la falta de apoyo institucional (Consejo de Rectoras y Rectores de las Universidades Chilenas [CRUCH], 2024). Esta falta de apoyo no necesariamente es por falta de incentivo por parte de la institución de educación superior, sino por falta de una herramienta que sea capaz de detectar a estudiantes susceptibles de presentar bajos rendimientos académicos. Debido a esto, es relevante identificar los factores académicos de ingreso, que puedan incidir en el rendimiento de los estudiantes de Educación General Básica, pues gracias a esto, se podrá crear un modelo de predicción, y a su vez, permitan diseñar estrategias de acompañamiento de dichos estudiantes, contribuyendo a mejorar la calidad formativa y reducir el riesgo de bajo rendimiento, o en el peor de los escenarios, deserción académica (Pérez y Soto Ardila, 2022). Esta propuesta, además de ser una herramienta útil para la Universidad de Concepción, es importante y necesaria, pues la diversidad de los estudiantes es cada vez mayor, y con ello, las nuevas dificultades que demandan dichas cohortes de estudiantes (Hurtado Palomino *et al.*, 2021).

Finalmente, en la Universidad de Concepción, la revisión documental no evidencia la existencia de estudios publicados ni de herramientas predictivas que utilicen antecedentes académicos de ingreso para anticipar el rendimiento académico en la carrera de Educación General Básica. Si bien la institución dispone de mecanismos de seguimiento académico y evaluaciones diagnósticas de ingreso, estos se orientan a la caracterización general de las cohortes y no a la construcción de modelos predictivos que permitan identificar tempranamente a estudiantes susceptibles de un bajo rendimiento. Esta ausencia de evidencia local refuerza la pertinencia y necesidad del presente estudio, en tanto contribuye a generar conocimiento contextualizado y a sentar las bases para el desarrollo futuro de estrategias de apoyo académico basadas en datos en la Universidad de Concepción.

Capítulo II: Marco Teórico.

En educación superior, el rendimiento académico es un tema complejo e interesante de estudiar (Schmidt *et al.*, 2023), pues, si bien hay factores externos al estudiante que condicionan su rendimiento académico, existen factores intrínsecos del propio estudiante que inciden fuertemente en su rendimiento académico en educación superior, como el historial académico durante su formación escolar previa al proceso de admisión universitaria (Herrera Rivera y Arancibia Carvajal, 2022).

Un análisis de esta relación es importante para comprender las barreras que enfrentan los estudiantes, especialmente durante sus primeros años de formación, y, así, sentar las bases de estrategias que promuevan el apoyo, y posteriormente, la aprobación de las asignaturas que requieren un alto compromiso personal y cognitivo.

2.1. Factores académicos que inciden en el rendimiento en la educación superior

El rendimiento académico en la educación superior ha sido objeto de amplio estudio, ya que constituye un indicador de la calidad del aprendizaje y de la efectividad de los procesos formativos. Su análisis busca comprender cómo las características propias del estudiante, sus experiencias previas y las condiciones institucionales que influyen en su desempeño a lo largo de su trayectoria universitaria.

Una de las aproximaciones más influyentes para abordar este fenómeno es el Modelo 3P de Biggs (2001), el cual concibe el rendimiento académico a partir de tres dimensiones: las condiciones de entrada, los procesos de aprendizaje y los productos o resultados. Las condiciones de entrada incluyen los conocimientos, habilidades y experiencias con las que el estudiante inicia su formación, mientras que los procesos reflejan las estrategias y dinámicas de aprendizaje que desarrolla durante su paso por la universidad. Desde esta perspectiva, los puntajes de ingreso y las evaluaciones diagnósticas representan variables de entrada relevantes que permiten anticipar el rendimiento académico en asignaturas universitarias.

De acuerdo con Hernández Knipp y Zúñiga Lalanne (2019), el rendimiento académico se construye a partir de la interacción entre elementos personales y contextuales, los cuales se manifiestan en las calificaciones obtenidas durante la formación universitaria. En esta misma línea, Cobo Rendón *et*

al. (2022) identifican que las habilidades cognitivas desarrolladas antes del ingreso a la educación superior determinan la capacidad de los estudiantes para adaptarse al ritmo académico y enfrentar evaluaciones más complejas, lo que reafirma la relevancia de las condiciones de entrada descritas por Biggs.

Asimismo, Román González *et al.* (2020) distinguen dos dimensiones esenciales para el aprendizaje disciplinar: por un lado, están las habilidades cognitivas, vinculadas con la interpretación y aplicación de la información, por otro lado, están las habilidades de estudio, relacionadas con la organización y gestión del aprendizaje. Ambas suelen medirse mediante puntajes obtenidos en pruebas estandarizadas o diagnósticos institucionales, que funcionan como indicadores comparables del nivel de preparación inicial del estudiante.

En concordancia con lo anterior, Jérez Yáñez *et al.* (2018) sostienen que las trayectorias formativas previas condicionan la forma en que los estudiantes procesan los nuevos contenidos, especialmente en áreas que demandan razonamiento abstracto y aplicación de conceptos. De la misma manera, Vega Vargas (2025) enfatiza la importancia de los recursos institucionales de apoyo en la educación superior, los cuales contribuyen a enfrentar de mejor manera las exigencias de los primeros años de formación, reforzando así la conexión entre preparación previa y acompañamiento institucional.

Del mismo modo, Cervini (2018) interpreta el rendimiento académico como una interacción de variables, que integra tanto los aprendizajes previos como las oportunidades que ofrece el sistema universitario para el desarrollo disciplinar y cognitivo del estudiante. Hernández Yépez *et al.* (2022) reportan, a partir de su estudio, que los factores académicos asociados a la formación previa presentan una mayor incidencia en el rendimiento académico universitario en comparación con otras variables de carácter cualitativo, como las socioemocionales o sociodemográficas, las cuales muestran una relación más débil o indirecta con el rendimiento académico.

El acompañamiento a estudiantes universitarios en sus primeros años de formación, ha evidenciado que la formación escolar previa, en muchos casos caracterizada por un aprendizaje superficial, condiciona el desempeño académico posterior (Ruiz y Torres, 2020). Estos mismos autores, destacan la necesidad de fortalecer los apoyos institucionales que permitan mitigar las brechas de conocimiento y facilitar la adaptación al contexto universitario.

En conclusión, las investigaciones coinciden en que el rendimiento académico no puede entenderse como un resultado aislado, sino como el reflejo de un proceso formativo que combina experiencias acumuladas y exigencias propias de la educación superior. Variables como el dominio matemático, la comprensión lectora y la capacidad de resolución de problemas se consolidan como elementos determinantes para anticipar el rendimiento académico en asignaturas de alta complejidad. Estas perspectivas constituyen la base de esta investigación, al evidenciar que los factores académicos de ingreso pueden utilizarse como herramientas predictivas para identificar tempranamente a estudiantes susceptibles de un bajo rendimiento académico.

2.2. Factores académicos que inciden en el rendimiento académico en las asignaturas de Física y Química

Las asignaturas de Física y Química en la educación superior presentan características propias que requieren un nivel de razonamiento y comprensión conceptual más elevado que otras áreas del conocimiento. Estas disciplinas demandan la aplicación de habilidades matemáticas, interpretación de modelos y comprensión de representaciones simbólicas, lo que incrementa su nivel de dificultad para los estudiantes que no cuentan con una base escolar robusta en estas áreas. En el caso de la asignatura de Física, Barahona U. (2014) describe esta asignatura como una de las más complejas dentro de los primeros años del ciclo universitario, debido a su énfasis en el razonamiento lógico-matemático y en la resolución de problemas que implican modelizar fenómenos naturales. Del mismo modo, Guillart *et al.* (2020), la describen como compleja para los estudiantes de educación superior, debido a la falta de un desarrollo de las habilidades del pensamiento científico investigativo al momento de ingresar a la universidad, especialmente, las habilidades relacionadas con la búsqueda de información científica.

De manera similar, la asignatura de Química exige interpretar información abstracta y comprender procesos que combinan lo conceptual con lo cuantitativo. Barahona Ibarra *et al.* (2024) sostienen que el dominio previo de habilidades numéricas y el entendimiento de conceptos científicos básicos influyen de manera significativa en el desempeño de los estudiantes. La comprensión de los lenguajes simbólicos propios y la capacidad de realizar cálculos específicos requieren que los estudiantes ingresen con una preparación adecuada para enfrentar los contenidos universitarios. Ahora bien, focalizándose en una carrera de pedagogía, el panorama no es muy diferente. Agila Cuenca (2020) señala que, en la carrera de Pedagogía en Química, la mayor incidencia condicionante del rendimiento académico es en base a

formación preuniversitaria del estudiante, seguida de una carencia de condiciones institucionales, como la falta de acompañamientos por parte de la institución de educación superior.

Por ello, Garzon *et. al.* (2010) se centran en los factores académicos de ingreso como predictor del rendimiento académico en las asignaturas de Física y Química, en la educación superior. De este modo, destacan la incidencia de las competencias académicas de los estudiantes, traducidas en resultados académicos previos al proceso de admisión universitaria. Por consiguiente, los autores de este estudio enfatizan en la importancia de fortalecer las estrategias de apoyo y orientación continua por parte de la universidad, para enfrentar las deficiencias de formación escolar y la dificultad de la transición académica, donde las instituciones de educación superior deben asumir un rol activo en la nivelación de competencias, especialmente en las áreas científicas, donde las brechas de conocimiento suelen ser más amplias.

Por lo tanto, los factores académicos de ingreso adquieren especial relevancia para comprender el rendimiento en ambas disciplinas. Su análisis resulta importante para anticipar posibles dificultades y comprender de manera más precisa el desempeño académico en estas áreas dentro de la carrera de Educación General Básica.

2.3. Modelos predictivos del rendimiento académico

Los modelos predictivos se han consolidado como herramientas útiles y funcionales para identificar patrones en estudiantes con riesgo de un bajo rendimiento académico. Díaz Landa *et al.* (2021) y Mc Coll Calvo *et al.* (2021) señalan que estos modelos operan a partir de un análisis integrado de variables, combinando distintos tipos de datos académicos (pruebas estandarizadas, resultados académicos de ingreso y rendimiento académico de asignaturas, etc.). Esta integración, como señalan González y Muñoz (2021), facilita la detección de estudiantes susceptibles, lo que permite a las instituciones diseñar estrategias de apoyo oportunas.

La precisión de los modelos no solo depende de las variables empleadas, sino también de la rigurosidad con que se valida su funcionamiento (Castrillón, Sarache y Ruiz Herrera, 2020). En este sentido, el estudio de Núñez Villalobos (2025), destaca que técnicas como la regresión logística permiten identificar con alta exactitud a los estudiantes en riesgo, lo que facilita la implementación de intervenciones personalizadas. Un paso fundamental en este proceso es la validación del modelo, ya que

garantiza que no solo funcione con los datos originales, sino que pueda generalizar su capacidad predictiva a nuevos grupos de estudiantes. Para ello, Henríquez Cabezas y Vargas Escobar (2022) utilizan la curva *Receiver Operating Characteristic (ROC)* y el área bajo la curva (AUC), como métricas de validación, pues permiten evaluar la capacidad general del modelo para diferenciar entre estudiantes con alto y bajo rendimiento académico.

Para garantizar la fiabilidad del modelo, se emplean métricas específicas y técnicas de validación (Bello *et al.*, 2020; Mella Norambuena *et al.*, 2022). La validación cruzada (como la validación *k-fold*) evalúa la exactitud del modelo en diferentes subconjuntos de datos, lo que reduce el riesgo de que el modelo esté sobre ajustado. Otra métrica clave es la sensibilidad, que mide la capacidad del modelo para identificar correctamente a los estudiantes en riesgo y la especificidad, que evalúa su capacidad para identificar a quienes son o no susceptibles a tener un bajo rendimiento académico (Del Carpio Mendoza, 2024).

Para la aplicación de un modelo predictivo, se requiere incorporar de manera precisa las condiciones iniciales del estudiante como puntajes en las pruebas de admisión, diagnósticas, etc., para posteriormente, configurarlos como predictores significativos del rendimiento académico en una asignatura objetivo. Peralta De La Vega (2005), a través de un estudio estadístico, estableció una influencia directa entre los factores académicos de ingreso y el rendimiento académico posterior en las asignaturas del grupo de las Ciencias Naturales. Los resultados de su análisis evidenciaron una correlación directa entre ambas variables, demostrando que la preparación previa condiciona el rendimiento académico universitario. Adicionalmente, identificó que la asignatura de Física presenta la mayor tasa de reprobación, con más del 50% de los estudiantes no aprobados, seguida por Química con un 45% aproximadamente, y Biología con un 41% aproximadamente.

Estos hallazgos refuerzan la necesidad de incluir los antecedentes académicos de ingreso como una variable clave dentro de los modelos predictivos del rendimiento académico, ya que su análisis permite identificar algunas posibles causas del bajo rendimiento académico en las asignaturas de Física y Química. En consecuencia, para que estos antecedentes académicos de ingreso puedan ser incorporados de manera efectiva en un modelo predictivo, es necesario traducirlos desde un nivel conceptual a un nivel operativo. Esto implica definir con precisión qué variables representarán empíricamente dichos factores

y cómo estas se relacionan con el rendimiento académico observado. La operacionalización de las variables permite establecer un vínculo directo entre la teoría que sustenta el rendimiento académico y las técnicas estadísticas empleadas, asegurando que el modelo predictivo mantenga coherencia conceptual y una validez metodológica.

En la **tabla 1**, se muestra cómo cada concepto abstracto, como los antecedentes académicos de ingreso o el rendimiento académico, se transforma en indicadores medibles dentro de un modelo estadístico, asegurando coherencia entre el marco teórico, la metodología y el análisis de datos. De esta manera, se establece con claridad la forma en que las variables se integran al proceso de predicción del desempeño en Física y Química.

Tabla 1. Cuadro de operacionalización y correspondencia de constructos.

Constructo.	Variable.	Hipótesis.	Técnica Analítica.
<i>Factores académicos de entrada.</i>	Resultados prueba de ingreso (PSU, PDT, PAES), y pruebas de diagnóstico UdeC.	El rendimiento académico de ingreso es un predictor del rendimiento académico en las asignaturas de Física y Química en su segundo año universitario.	Árboles de decisión / Validación cruzada (k-fold).
<i>Rendimiento académico.</i>	Promedio de notas en las asignaturas de Física y Química.	El rendimiento académico en las asignaturas de Física y Química se asocia con las notas en las evaluaciones de ingreso (pruebas de ingreso y evaluaciones diagnósticas) y de la dependencia del establecimiento educacional de egreso.	Variable dependiente en los modelos estadísticos.

Capítulo III: Marco Metodológico.

La presente investigación se fundamenta en una perspectiva epistemológica que se distingue por su enfoque cuantitativo y la búsqueda de leyes generales libre de la subjetividad (Acosta Faneite, S. F., 2023). Éste asume la existencia de una realidad externa y objetiva que puede medirse, permitiendo al investigador analizar relaciones empíricas entre variables a través de la observación sistemática y la recopilación de datos (Ávila Morales *et al.*, 2014).

En este estudio, el paradigma positivista se orienta a examinar las relaciones y asociaciones existentes entre variables académicas, así como su capacidad predictiva respecto del rendimiento académico. Mediante la cuantificación y el análisis de datos numéricos, se busca identificar patrones y vínculos estadísticamente significativos que contribuyan a la comprensión del rendimiento de los estudiantes, sin establecer relaciones causales directas.

3.1. Tipo o Enfoque de la Investigación

El presente estudio se enmarca en un enfoque cuantitativo, con un diseño no experimental, transversal y retrospectivo, de carácter exploratorio, predictivo y correlacional. Este enfoque metodológico es idóneo para analizar estadísticamente la relación entre diversas variables académicas y su capacidad para predecir el rendimiento académico de los estudiantes en asignaturas de ciencias, específicamente en Física y Química (Hernández Sampieri *et al.*, 2014), a partir del análisis de datos académicos históricos correspondientes a las cohortes 2018-2024, los cuales son examinados en un único momento, sin seguimiento longitudinal. Estas asignaturas son cursadas en el segundo año de educación superior en la carrera de Educación General Básica.

3.2. Diseño de Investigación

En el estudio, al ser de tipo no experimental, no se manipularán las variables y la información completa se extraerá simultáneamente entre sí. Por consiguiente, es de carácter correlacional y predictivo (Morel, A., 2025), pues se busca identificar relaciones significativas entre variables independientes como la dependencia y modalidad del establecimiento educacional de origen (particular subvencionado, particular pagado o municipal). También se considerarán las variables de entrada de los estudiantes, como los resultados de pruebas de selección universitaria (PSU, PDT y PAES) y pruebas de diagnóstico de la Universidad de Concepción de las cuatro asignaturas fundamentales (Ministerio de Educación - Centro

de Estudios, 2023). Además, es predictivo, al pretender desarrollar un modelo estadístico que permita estimar el rendimiento académico a partir del promedio ponderado semestral en las asignaturas de Física y Química.

3.3. Dimensión Temporal

La presente investigación es de corte transversal, ya que los datos se recopilarán en un único momento para examinar el rendimiento académico de los estudiantes de la carrera de Educación General Básica. Se considerarán las cohortes recientes de estudiantes (2018-2024). Este enfoque permite identificar patrones actuales y correlaciones entre las variables estudiadas, como los antecedentes académicos de entrada y el rendimiento de los estudiantes. El objetivo es analizar estas relaciones en un solo punto, sin realizar un seguimiento longitudinal para observar su evolución a lo largo del tiempo.

3.4. Alcance de la Investigación

Esta investigación, al ser de tipo descriptivo, exploratorio, correlacional y predictivo, se centra en caracterizar las condiciones académicas de ingreso y analizar su relación con el rendimiento al año posterior a éste. A partir de ello, se desarrollará un modelo predictivo basado en RF, SVM, XGBoost, etc. (Fernández Casal *et al.*, 2024), con el fin de desarrollar un modelo que permitirá anticipar el rendimiento académico en las asignaturas de Física y Química en segundo año de la carrera.

3.5. Población

La población de investigación estará compuesta por la totalidad de los estudiantes que ingresaron a la carrera de Educación General Básica y que correspondan a las cohortes 2018-2019 y 2021-2024, excluyéndose la cohorte 2020 debido a la ausencia de resultados de pruebas diagnósticas institucionales.

Se estima que esta población asciende a un total de aproximadamente 200 estudiantes, distribuidos en seis cohortes efectivamente analizadas.

3.6. Participantes del Estudio

Los participantes de esta investigación son estudiantes que han ingresado a primer año de la carrera de Educación General Básica de la Universidad de Concepción Campus Los Ángeles, correspondientes a las cohortes 2018-2019 y 2021-2024, considerando exclusivamente a aquellos con información académica completa disponible para las variables de interés. En este sentido, la muestra

corresponde a una muestra no probabilística por disponibilidad, determinada por el acceso a registros institucionales válidos y completos.

Se consideraron sólo ingresados por la vía regular de admisión y vía de ingreso especial PACE (Programa de Acceso a la Educación Superior), o vía especial propedéutico. Se excluirán del estudio aquellos estudiantes que hayan realizado cambio de carrera, poseedores de un título profesional y aquellos provenientes del programa de articulación de la Universidad de Concepción.

Estos criterios de selección permiten asegurar la homogeneidad de la muestra, garantizando que los análisis predictivos del rendimiento académico se realicen sobre un grupo representativo de estudiantes con trayectorias de ingreso comparables. Así mismo, el tamaño muestral final, queda condicionado por la disponibilidad de datos por cohorte y por categoría de ingreso. La exclusión de estudiantes con condiciones de admisión distintas busca evitar sesgos que puedan alterar la validez de los resultados, permitiendo una interpretación más precisa de los factores académicos que inciden en el rendimiento de los estudiantes en las asignaturas de Física y Química.

3.7. Unidad de Análisis

La unidad de análisis de este estudio son los estudiantes de segundo año que ingresaron a la carrera de Educación General Básica en la Universidad de Concepción, campus Los Ángeles, entre los años 2018 y 2024, y que cursaron las asignaturas de Física y Química. El foco de la investigación es examinar la relación entre los resultados académicos de ingreso de cada estudiante con su rendimiento académico en las asignaturas de Física y Química.

3.8. Recolección de Datos

Se utilizará la técnica de revisión documental como instrumento principal (Martínez-Corona *et al.*, 2023), considerando:

- Registros académicos, es decir, actas de notas de las asignaturas de Física y Química, con su respectiva cuantificación.
- Puntajes de las pruebas de selección universitaria de matemática, lenguaje y ciencias.
- Resultados de pruebas de diagnóstico UdeC de primer año en las asignaturas de matemáticas, lenguaje, historia y ciencias naturales.
- Dependencia y modalidad del establecimiento de egreso de la educación escolar.

3.9. Técnica de Análisis de Datos

El análisis de los datos se realizará en dos fases. La primera fase comprende un análisis descriptivo y exploratorio. En esta etapa se calcularán medidas de tendencia central (dispersión y distribución para cada variable) estadísticos como media, mediana, desviación estándar y percentiles. Además, se comprobará la presencia de datos atípicos, datos faltantes y se determinará el tipo de distribución que tienen los datos.

Previo al entrenamiento de los modelos predictivos, se realizó una serie de ajustes metodológicos fundamentados en la estructura de los datos:

- Simplificación de la estructura categórica:

En primer lugar, se creó la variable categórica fusionada DEPENDENCIA_MODALIDAD, con el objetivo de reducir la fragmentación categórica o representación de grupos y asegurar una representación adecuada de los distintos contextos educativos. Esta variable se utilizó exclusivamente como predictor estructural y fue codificada mediante *one-hot encoding*.

- Normalización de escalas:

Dado que las variables predictoras poseen escalas heterogéneas, se aplicó un proceso de normalización:

- Variables PTJE_* divididas por 1000.
- Variables PD_*, así como las variables objetivo FIS y QUI, divididas por 7.

Esta decisión permitió eliminar efectos artificiales del rango numérico, favoreciendo la estabilidad de modelos sensibles a escala y distancia.

- Tratamiento de valores ausentes:

El análisis exploratorio mostró una alta proporción de valores ausentes en PTJE_MAT2 y PTJE_CS, asociadas a que dichas pruebas tienen carácter optativo dentro del proceso de admisión universitaria y, por tanto, no son rendidas por la totalidad de los estudiantes. Adicionalmente, estas variables, presentaron una baja correlación con las variables objetivo. En este contexto específico, y con el fin de evitar la reducción significativa del tamaño muestral y la introducción de sesgos derivados de la imputación por la mediana, aplicada únicamente a los predictores dentro del proceso de entrenamiento.

- Selección de modelos robustos:

Considerando la no normalidad, la presencia de *outliers* reales y la colinealidad entre predictores, se optó por modelos robustos y no paramétricos:

- Random Forest (RF)

- Support Vector Regression (SVM)
- XGBoost (XGB)
- Regresión Ridge
- Regresión Lasso

- Evaluación por escenarios de predictores:

Para cada variable objetivo se evaluaron escenarios diferenciados, los que incluyen puntajes pruebas estandarizadas (PTJE), notas pruebas diagnósticas (PD) y la combinación de ambos conjuntos (PTJE + PD). De esta forma, el conjunto de variables predictoras para FIS y QUI, quedó de la siguiente forma:

- FIS: PTJE, PD y MIXTO (PTJE + PD)
- QUI: PTJE + FIS, PD + FIS y MIXTO + FIS

En el caso de la asignatura de Química, la variable FIS se justifica debido a que la asignatura de Física es cursada previamente dentro del plan de estudios, constituyendo un antecedente académico inmediato y disponible al momento de modelar el rendimiento en Química. En este sentido, FIS no representa una relación de temporalidad inversa, sino una variable académica previa que refleja competencias científicas y cuantitativas ya desarrolladas por el estudiante.

Para evaluar el desempeño del modelo, adicionalmente, se emplearán métricas RMS, MAE y R2. Todo el análisis se llevará a cabo mediante software estadístico de código abierto, “*R-studio*” (Sánchez Jiménez, 2024).

3.10. Consideraciones Éticas

El presente estudio respetará los principios éticos establecidos por la Asociación Americana de Psicología (APA, 2020). Para garantizar la confidencialidad y el anonimato, se codificarán los datos recolectados, protegiendo así la privacidad de los estudiantes y asegurando en todo momento el uso ético de la información institucional.

Capítulo IV: Resultados.

4.1 Análisis estadístico preliminar

La base de datos original, suministrados por la Unidad de Registros Académicos UdeC L.A. se traspasó a una hoja de cálculo para su posterior limpieza y aplicación de filtros de exclusión. Se eliminaron los datos de aquellos estudiantes provenientes de cambios de carrera y programas especiales de ingreso. Además, se eliminaron aquellos estudiantes que no tienen evaluaciones diagnósticas, como la cohorte 2020 y también, aquellos que no tienen calificaciones en las asignaturas de Física y Química. Después de la aplicación de estos filtros, la tabla de datos se compone de 15 variables con 160 observaciones cada una. Las variables que se analizarán en este estudio se muestran en la **tabla 2**.

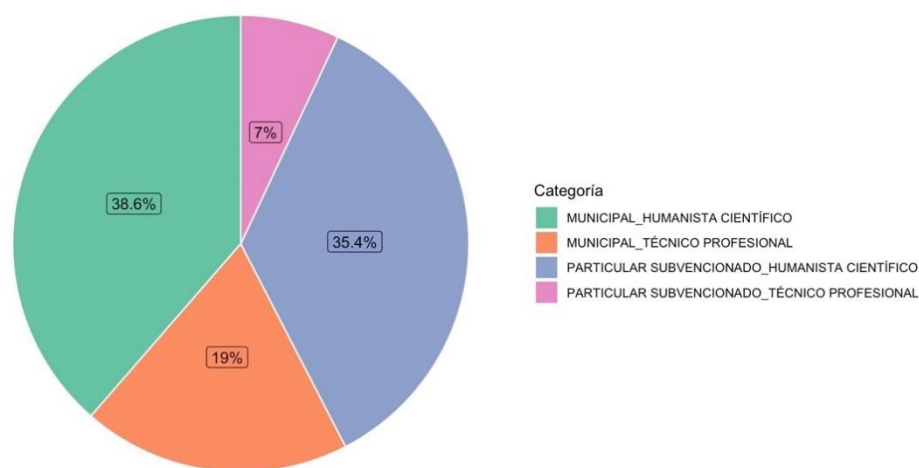
Tabla 2. Listado de variables y su respectiva descripción.

Nombre y/o abreviatura	Descripción
AÑO	Año de ingreso de cohorte
DEPENDENCIA	Dependencia administrativa establecimiento escolar
MODALIDAD	Tipo de educación impartida establecimiento escolar
PTJE_LEN	Puntaje en la prueba de Lenguaje y comunicación (obligatoria)
PTJE_MAT	Puntaje en la prueba de Matemática (obligatoria)
PTJE_MAT2	Puntaje en la prueba de Matemática 2 (optativa)
PTJE_CS	Puntaje en la prueba de ciencias (optativa)
PD_MAT	Nota prueba de diagnóstico de Matemática
PD_LEC	Nota prueba de diagnóstico de Lectura
PD_FIS	Nota prueba de diagnóstico de Física
PD_BIO	Nota prueba de diagnóstico de Biología
PD_QUI	Nota prueba de diagnóstico de Química
FIS	Nota asignatura de Física
QUI	Nota asignatura de Química

Las variables que se emplearon en este estudio comprenden algunas de tipo categóricas como año de ingreso, dependencia del establecimiento educacional de origen y la modalidad del tipo de educación impartida. Las otras variables son del tipo cuantitativa y corresponde a puntajes obtenidos en las pruebas de selección a la educación superior, pruebas diagnósticas aplicadas al ingreso a universidad y también, las calificaciones obtenidas en las asignaturas de Física y Química.

Una vez que los datos fueron filtrados se realizó un análisis estadístico descriptivo con el propósito de conocer la estructura y distribución de los datos. En la **figura 1**, se muestran la distribución de estudiantes según dependencia y modalidad de los establecimientos educacionales de origen. Como se observa en esta figura, el 58% de estudiantes provienen de la educación municipal y el 42% restante, de la educación particular subvencionada. En cuanto al tipo de modalidad el 74 % de los estudiantes provienen de un establecimiento científico humanistas y un 26 % de establecimientos técnico profesional.

Figura 1: Distribución de estudiantes según tipo de dependencia y modalidad de los establecimientos educacionales de origen.



En la **tabla 3**, se resumen los parámetros estadísticos y sus respectivos valores de las variables incluidas en este estudio.

Tabla 3. Parámetros estadísticos de las variables académicas.

	Variable										
	PTJE LENG	PTJE MAT	PTJE MAT2	PTJE CS	PD MAT	PD LEC	PD FIS	PD QUI	PD Blo	FIS	QUI
Min.	388	294.	313	268	1.000	1.800	1.2	1.0	1.0	2.6	3.0
1^{er} Q	559	507	399	454	2.300	3.000	2.0	2.4	2.5	4.2	4.2
Mediana	608	565	430	502	2.700	3.400	2.4	2.7	2.9	4.7	4.5
Media	619	560.2	431	499.9	2.949	3.426	2.488	2.8	3.027	4.799	4.541
3^{er} Q	674	607.5	476	562	3.500	3.900	2.9	3.2	3.500	5.5	4.8
Max.	853	769	522	682	6.400	5.800	4.9	4.4	6.300	6.5	5.9

Los puntajes obtenidos en las pruebas de ingreso señalan en el caso de Lenguaje, una distribución de puntajes bastante amplia, con una ligera cola hacia la derecha, es decir, alumnos con puntajes más altos, esta distribución evidencia la presencia de estudiantes con resultados elevados en esta prueba. En

cuanto a la prueba de matemática, el comportamiento es similar a la prueba de lenguaje, pero con una distribución más simétrica.

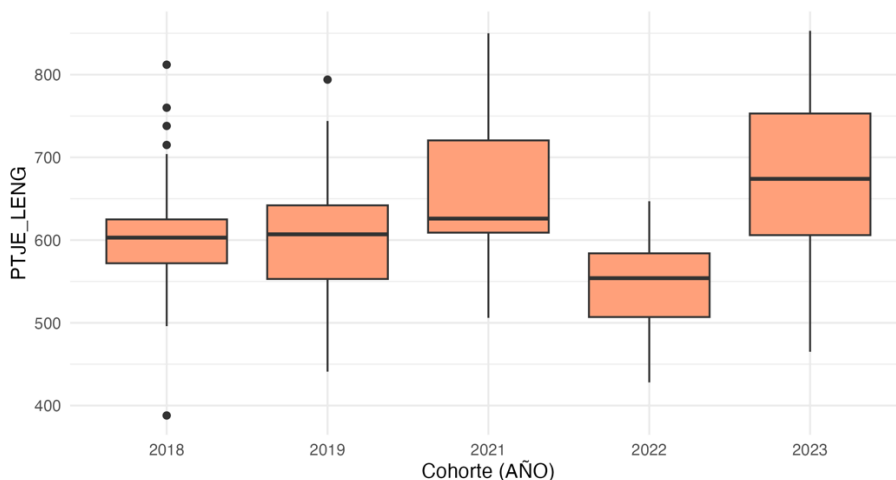
Por su parte, en la prueba de matemática-2 (PTJE_MAT2), los puntajes están distribuidos en un rango más estrecho, además estos muestran una distribución más equilibrada, ya que el valor de la media y mediana son muy similares. En el caso de la carrera de Educación General Básica, no es requisito de ingreso rendir esta prueba, lo que se traduce en una menor cantidad de observaciones disponibles para esta variable. Por otro lado, la prueba de ciencias presenta un comportamiento muy similar a la prueba de matemática 2, con una distribución bastante simétrica. En conjunto, los puntajes de PTJE_MAT2 y ciencias (PTJE_CS), permiten distinguir el desempeño de los estudiantes en evaluaciones asociadas al ámbito científico matemático.

En cuanto a las evaluaciones diagnósticas, todas presentan valores que indican un bajo desempeño, con distribuciones casi simétricas. Estos resultados muestran que una proporción importante de los estudiantes obtiene calificaciones en los tramos inferiores de la escala y, en el caso de las pruebas diagnósticas de Física y Química, se observan promedios bajos en relación con el máximo posible de la escala. Las pruebas diagnósticas de Matemática y Biología muestran que existe un pequeño grupo de estudiantes con calificaciones considerablemente más altas, los cuales podrían identificarse como valores atípicos, pero que corresponden a observaciones válidas dentro del conjunto de datos.

Finalmente, en cuanto a las asignaturas de Física y Química, se observa que, en el primer caso, la distribución presenta un ligero sesgo hacia la derecha, junto con un rango intercuartílico amplio. Por su parte, la asignatura de Química presenta una distribución más simétrica y un rango intercuartílico más estrecho en comparación con Física. A partir de los valores descritos en la **tabla 3**, se observa que las calificaciones finales en ambas asignaturas presentan distintos niveles de dispersión, siendo mayor en Física que en Química. Asimismo, el perfil de ingreso de los estudiantes se caracteriza por una marcada heterogeneidad, reflejada en la amplitud de los puntajes en las pruebas de Matemática y Lenguaje. Las pruebas diagnósticas, por su parte, concentran una alta proporción de calificaciones en los tramos inferiores de la escala, lo que evidencia diferencias importantes en los niveles de desempeño inicial de los estudiantes al momento de ingresar a la carrera.

A partir de la tabla de los parámetros estadísticos, se han construido una serie de representaciones gráficas que proporcionan antecedentes complementarios a lo ya expuestos anteriormente. Las **figuras 2 y 3**, muestran los *boxplot* de los puntajes obtenidos en las pruebas de lenguaje y matemática para el ingreso a la educación superior según la cohorte de ingreso.

Figura 2: *Boxplot de los puntajes en las pruebas de selección universitaria de lenguaje por cohorte de ingreso.*

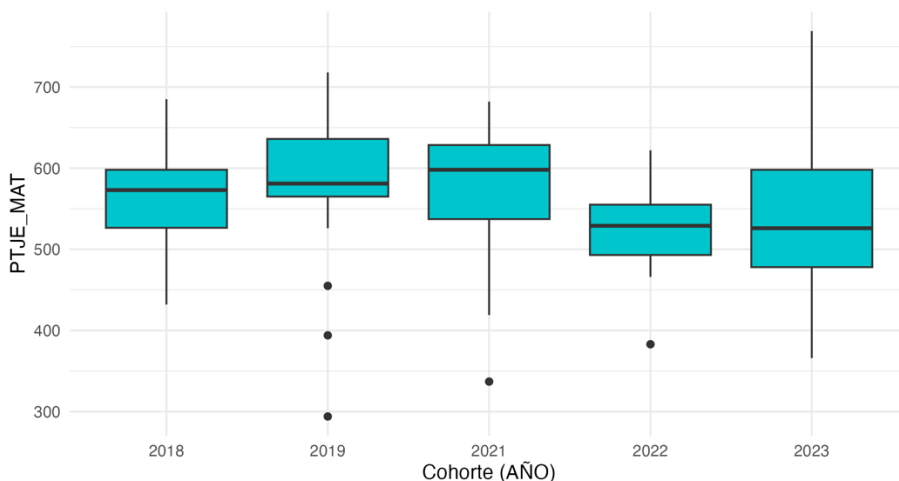


En el caso de los puntajes de la prueba de lenguaje (PTJE LENG), los *boxplots* muestran variabilidad importante entre cohortes, con medianas que fluctúan entre valores cercanos a 550 y 620 puntos. Las cohortes 2021 y 2023 presentan los puntajes más altos, reflejando un perfil de ingreso con mayor dominio en competencias lectoras. En contraste, la cohorte 2022 registra los puntajes más bajos en lenguaje, acompañada de una dispersión moderada. Un aspecto destacable, es la presencia de valores atípicos en varias cohortes, especialmente en 2018 y 2019, asociados a estudiantes con desempeños excepcionalmente altos. Esto sugiere la coexistencia de perfiles académicos muy diversos al interior de cada generación que ingresa a la carrera.

Con respecto a los puntajes de matemática (PTJE MAT), se observa un patrón similar al de la prueba de lenguaje, aunque con algunas particularidades. Las cohortes 2019 y 2021 presentan medianas más altas en comparación con las demás cohortes, mientras que la cohorte 2022 exhibe los valores más bajos tanto en la mediana como en el rango intercuartílico. La cohorte 2023 presenta una distribución más amplia, con mayor dispersión de los puntajes y presencia de valores tanto bajos como altos. Asimismo, se

identifican valores atípicos inferiores en algunas cohortes, principalmente en 2019 y 2021, los cuales se sitúan por debajo del rango intercuartílico y corresponden a observaciones individuales que se apartan del comportamiento central de la distribución. Ambos resultados revelan que las cohortes presentan perfiles académicos diferenciados al ingreso, tanto en Lenguaje como en Matemática.

Figura 3: *Boxplot de los puntajes en las pruebas de selección universitaria de matemática por cohorte de ingreso.*



La **figura 4**, muestra el *boxplot* de las calificaciones obtenidas en las asignaturas de Física y Química. Se observa que la asignatura de Física tiene mayor dispersión que la asignatura de Química, debido al mayor rango intercuartílico (IQR). Esto indica una mayor variabilidad en las calificaciones de la asignatura en comparación con Química. En el caso de la asignatura de Química, el IQR es más estrecho, lo que refleja una menor dispersión de las calificaciones. En ambas asignaturas el promedio es muy similar, observándose valores centrales comparables entre ambas, pero en la asignatura de Física existen grupos de estudiantes que presentan promedios mucho mayores que en Química. Por otro lado, la asignatura de Física no presenta valores anómalos, en cambio Química presenta valores mayores y menores que el IQR. La diferencia entre la variabilidad observada entre ambas asignaturas permite describir distribuciones con características distintas, sin embargo, el análisis de sus implicancias se aborda en apartados posteriores.

Figura 4: Boxplot de las calificaciones de física y química, corte 2018-2024.

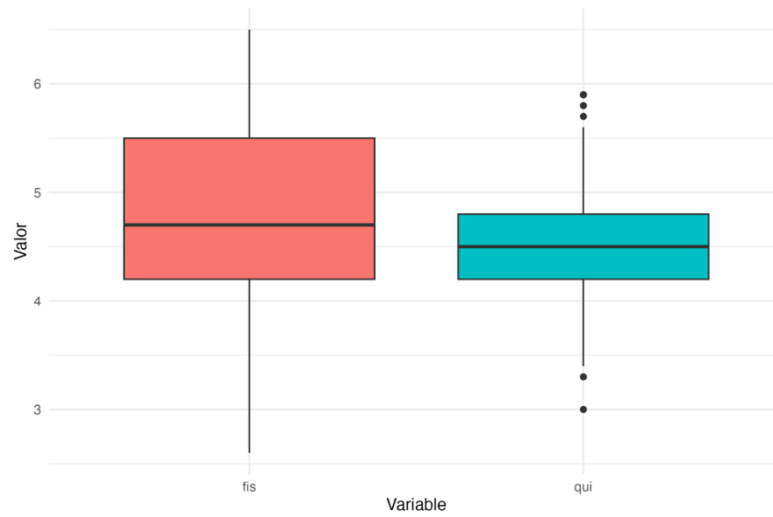
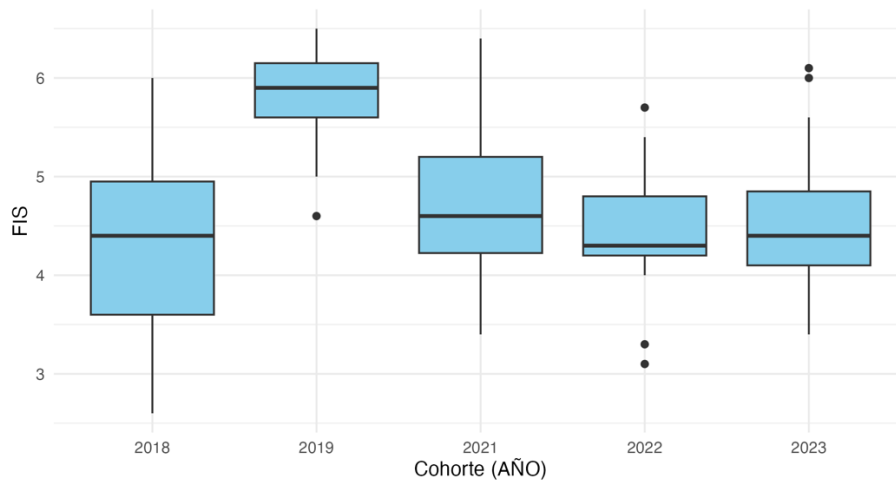


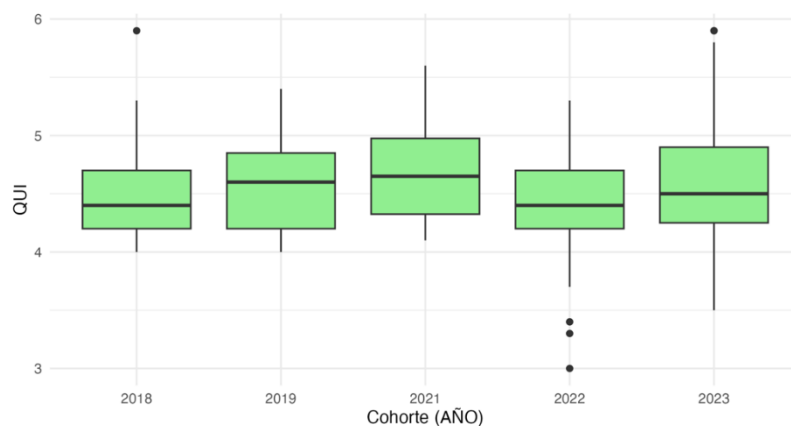
Figura 5: Boxplot de las calificaciones en física por cohorte.



La **figura 5**, muestra los *boxplots* de Física por cohorte. En esta serie destaca la cohorte 2019 que tiene una mediana muy superior al resto de las cohortes (≈ 6.0), con una IQR estrecho, sin observarse valores bajos dentro de su distribución. En el caso de la cohorte 2021, la mediana se ubica entre 4.7 y 4.8, pero con una mayor variabilidad de calificaciones. Las cohortes 2018 y 2022 muestran medianas más bajas. Además, se comprueba la presencia de datos atípicos en algunas de las cohortes, pero en un número limitado. Las diferencias observadas en las medianas y en la dispersión entre cohortes evidencian variabilidad en el rendimiento de la asignatura de Física a lo largo de los especialmente en el año 2019.

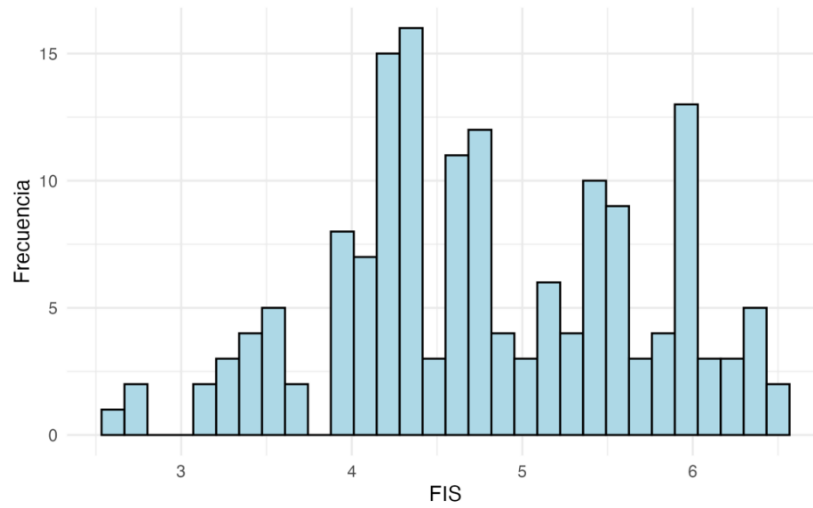
En el caso de la asignatura de Química (**figura 6**), las calificaciones por cohortes son más homogéneas que en Física, ya que las medianas fluctúan entre 4.4 y 4.7. La cohorte 2021 presenta un valor de mediana mayor, aunque las dispersiones entre cohortes son. Además, se observan valores *outliers* alrededor de 3.0 en la cohorte 2022. En términos descriptivos, las distribuciones de Química muestran menor variabilidad entre cohortes en comparación con Física, sin que se observen diferencias marcadas en las medidas centrales.

Figura 6. *Boxplot de las calificaciones de química por cohorte.*



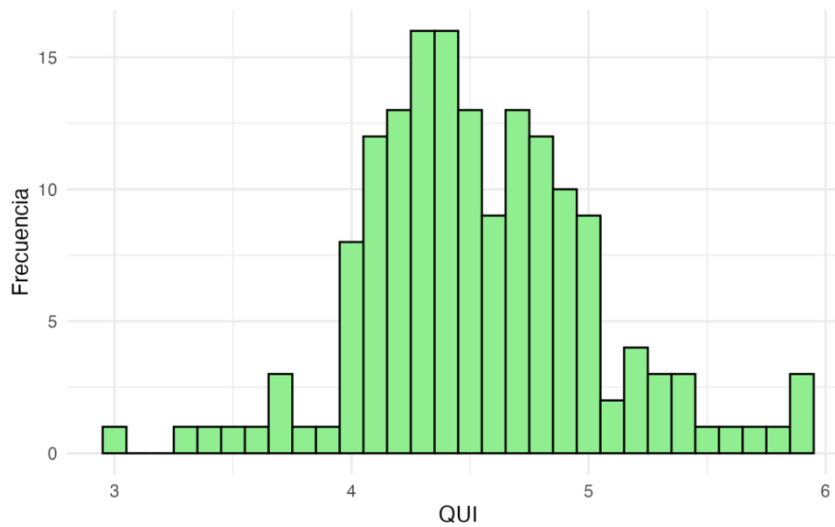
En la **figura 7**, se muestra el histograma de las calificaciones de Física. A primera vista, se puede observar que la distribución de calificaciones no es normal, ya que se aprecian varios máximos (forma multimodal). Se observan agrupaciones de estudiantes con calificaciones entre 4.0-4.3, otro entre 4.6-5,0 y otro con calificaciones mayores a 6.0. Este comportamiento multimodal, es coherente con el efecto de cohorte destacado previamente.

Figura 7. Histograma de las calificaciones de física desde el año 2018 hasta 2024.



A diferencia del histograma de las calificaciones de física, el histograma de química (**figura 8**), muestra una distribución más normal centrada entre 4.3-4.6, con un ligero sesgo hacia la derecha y con muy pocos valores muy altos o muy bajos.

Figura 8. Histograma de las calificaciones de química desde el año 2018 hasta 2024.

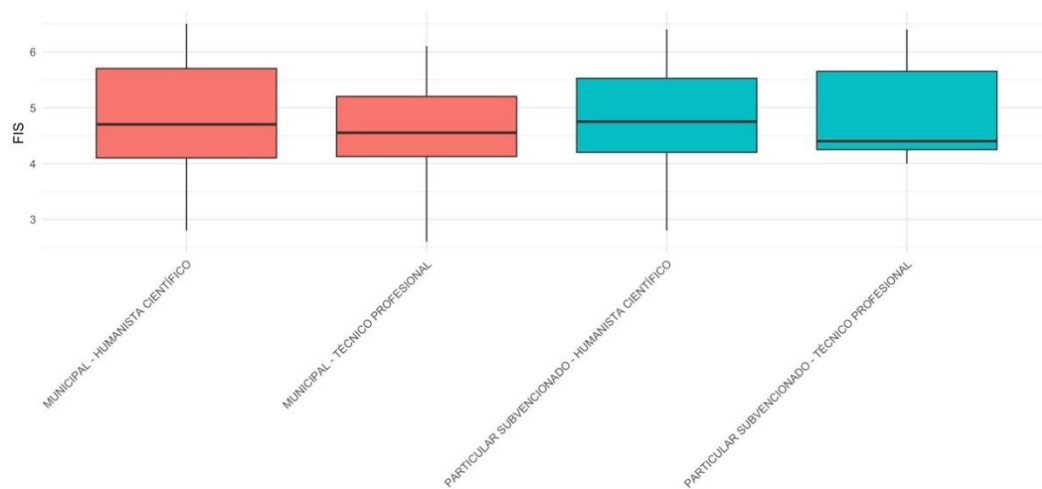


Sin embargo, en ambas situaciones, es necesario realizar los respectivos test de normalidad para así elegir el mejor algoritmo de regresión.

Con el propósito de explorar posibles diferencias en el rendimiento académico en función de características de la trayectoria escolar, se analizaron las calificaciones obtenidas en las asignaturas de Física (FIS) y Química (QUI) de acuerdo con la dependencia administrativa (Municipal o Particular Subvencionado) y la modalidad del establecimiento de origen (Humanista–Científico o Técnico–Profesional). Para ello, se construyeron *boxplots* que permiten comparar la distribución de notas en cada combinación Dependencia–Modalidad.

Los *boxplots* muestran que el rendimiento en Física (**figura 9**) presenta variaciones importantes según la dependencia y, en menor medida, según la modalidad del establecimiento educativo. En primer lugar, se observa que los estudiantes provenientes de establecimientos particulares subvencionados exhiben, en ambos tipos de modalidad, tendencias centrales levemente superiores a las observadas en los establecimientos municipales. No obstante, el grado de solapamiento entre cajas indica que estas diferencias no son marcadamente amplias.

Figura 9. *Boxplot de las calificaciones de física según dependencia y modalidad establecimiento educativo de origen.*

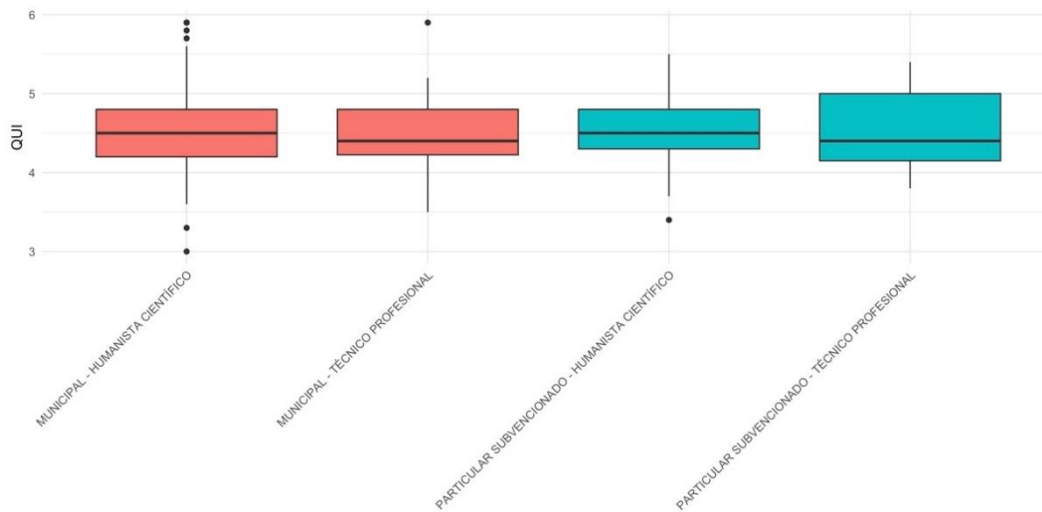


Dentro de cada dependencia, la comparación entre modalidades sugiere que los colegios Humanista–Científico tienden a concentrar distribuciones algo más altas que los Técnico–Profesionales, especialmente en el caso municipal. Aun así, los rangos intercuartílicos se mantienen relativamente similares, lo que sugiere que las diferencias observadas no necesariamente responden a un desplazamiento sistemático de todo el grupo, sino a variaciones puntuales dentro de la distribución.

Estos resultados muestran que no se observan diferencias marcadas en el rendimiento en Física según modalidad o dependencia, aunque, se registran medianas levemente superiores en estudiantes provenientes de establecimientos particulares subvencionados.

En el caso de la asignatura de Química (**figura 10**), se mantiene una tendencia similar a la observada en Física, aunque con algunas particularidades. En términos de dependencia, los estudiantes provenientes de establecimientos particulares-subvencionados muestran medianas marginalmente superiores y una concentración más acotada de sus valores, especialmente en la modalidad Técnico-Profesional. En contraste, los estudiantes del sector municipal presentan mayor dispersión, particularmente hacia los valores inferiores, junto con la presencia de valores atípicos en las categorías Humanista-Científico y Técnico-Profesional.

Figura 10. Boxplot de las calificaciones de química según dependencia y modalidad establecimiento educativo de origen.

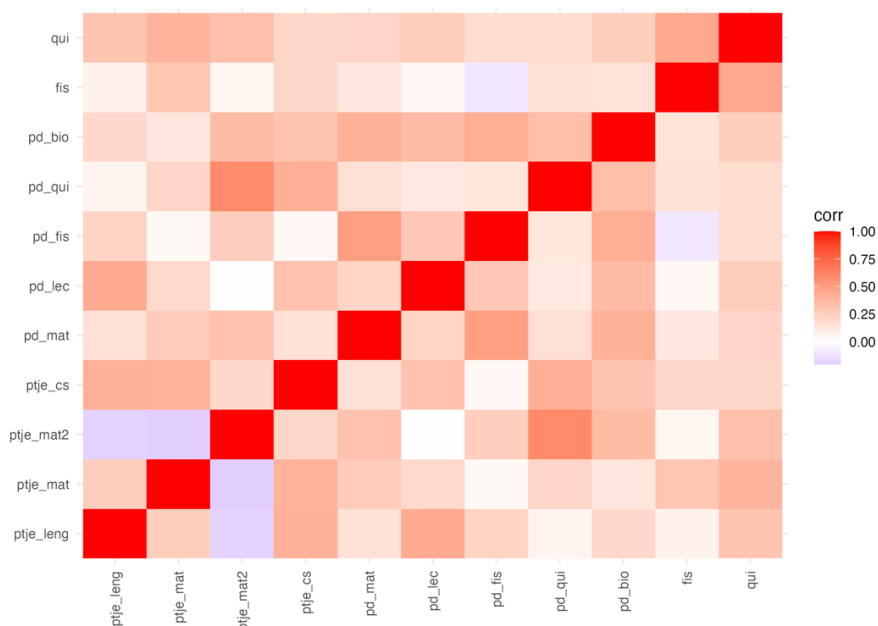


En relación con la modalidad, no se observa una diferencia marcada entre Humanista-Científico y Técnico-Profesional dentro de cada dependencia. Los valores de las medianas son relativamente próximos, lo que indica una distribución similar del rendimiento entre ambas modalidades en esta asignatura. Ahora, combinando ambos factores, se observa que las diferencias asociadas a la dependencia administrativa son levemente mayores que las asociadas a la modalidad, aunque las variaciones observadas siguen siendo moderadas no permiten establecer conclusiones inferenciales.

Los resultados de estos análisis constituyen insumos relevantes para la siguiente fase del análisis predictivo, dado que permiten contextualizar el comportamiento de los modelos y anticipar variables con capacidad discriminante limitada. En consecuencia, los resultados descriptivos presentados hasta el momento, contribuyen al entendimiento general de los datos y orientan la construcción de modelos no paramétricos en la siguiente etapa.

A partir de los análisis previos, es necesario establecer las posibles correlaciones entre las variables predictoras y las variables objetivo. En la **figura 11**, se muestra un mapa de calor (Heatmap) que relaciona las diversas variables académicas. Como se observa en este gráfico, existe una correlación moderada entre las asignaturas de Física y Química ($\sim 0.45 - 0.50$), lo que indica una asociación de magnitud media entre ambas variables, sin que ello implique una relación generalizable o causal. En el grupo de evaluaciones diagnósticas tienen correlaciones fuertes entre sí. Este mismo comportamiento es observado en el grupo de evaluaciones de ingreso a la educación superior. Finalmente, en cuanto a la asignatura de Física, se observan correlaciones más elevadas con PTJE_MAT, PD_MAT y PD_FIS. De forma análoga, la asignatura de Química presenta mayores correlaciones con las variables PTJE LENG, PD_LEC y PD_BIO. Estas asociaciones describen patrones de relaciones entre variables, sin establecer interpretaciones explicativas sobre los tipos de habilidades involucradas, las cuales serán abordadas más adelante.

Figura 11. Heatmap de correlaciones entre variables.



4.2 Pruebas de normalidad

Como parte del análisis exploratorio previo al modelamiento predictivo, se evaluó la normalidad de las variables cuantitativas incluidas en este estudio, mediante el test de *Shapiro-Wilk*, con el propósito de orientar la selección adecuada de técnicas estadísticas y decisiones metodológicas en las etapas posteriores del modelamiento predictivo. Los resultados del test se muestran en la **tabla 4**.

Tabla 4. Resultados del test de normalidad Shapiro-Wilk.

Variable	W	p-valor	Resultado
PTJE_LENG	0.9795	0.01749	No normal
PTJE_MAT	0.9915	0.46035	Normal
PTJE_MAT2	0.9573	0.41034	Normal
PTJE_CS	0.9608	0.01255	No normal
PD_MAT	0.9616	2.1e-4	No normal
PD_LEC	0.9883	0.20635	Normal
PD_FIS	964	3.8e-4	No normal
PD QUI	0.9859	0.10457	Normal
PD BIO	0.9678	8.7e-4	No normal
FIS	0.9751	0.00551	No normal
QUI	977	0.00904	No normal

Los resultados del test muestran que algunas de las variables presentan una distribución normal y otras no, lo que confirma la heterogeneidad de las poblaciones estudiadas y ya visualizadas previamente con los *boxplot* e histogramas. Estas características sugieren que estas variables podrían requerir de transformaciones del tipo Yeo-Johnson, logit o Boxcox, si se emplea una regresión lineal, no obstante, si se decide utilizar métodos más robustos como *Random Forest Regressor*, *Gradiente Boosting Regressor*, SVM con kernel RBF y Regresión lineal con transformación Yeo-Johnson. Para garantizar la comparabilidad y estabilidad en los modelos predictivos, es necesario evaluar la presencia de *outliers*, la pertinencia de la transformación de variables no normales y finalmente, un análisis de colinealidad para seleccionar el conjunto óptimo de predictores.

4.3 Análisis no paramétrico de las notas de Física y Química según contexto escolar

Dado que algunas de las variables predictoras no cumplen con el supuesto de normalidad, se realizó la prueba no paramétrica de *Kruskal-Wallis* para evaluar diferencias en las notas de Física (FIS)

y Química (QUI) según distintos factores contextuales, tales como la dependencia del establecimiento de origen, modalidad del colegio, combinación dependencia-modalidad y cohorte de ingreso (AÑO). Cuando corresponda, se aplicó el test *post hoc* de *Dunn* con corrección por comparaciones múltiples.

En primer lugar, se analizó la influencia de la variable dependencia del establecimiento escolar (municipal o particular subvencionado) en las calificaciones de FIS y QUI. Los resultados del test no evidenciaron diferencias estadísticamente significativas entre los grupos, ni para Física ($\chi^2 = 1,94$; $df = 3$; $p = 0,5842$) ni para Química ($\chi^2 = 2,05$; $df = 3$; $p = 0,562$). El análisis *post hoc* de *Dunn* confirmó este resultado, ya que todas las comparaciones por pares presentaron valores p ajustados superiores a 0,05, lo que indica que las medianas de las calificaciones de FIS y QUI son comparables entre los distintos tipos de dependencia.

En segundo lugar, se examinó el efecto de la modalidad del establecimiento (humanista-científico, técnico-profesional) como posible factor diferenciador. Al igual que en el caso anterior, la prueba de *Kruskal–Wallis* mostró ausencia de diferencias significativas en las notas de Física ($\chi^2 = 2,58$; $df = 2$; $p = 0,2757$) y Química ($\chi^2 = 0,29$; $df = 2$; $p = 0,8636$), lo que sugiere que la modalidad del colegio no se asocia de manera relevante con el rendimiento en estas asignaturas.

Con el propósito de explorar posibles efectos combinados, se creó una variable de grupo a partir de la interacción entre dependencia y modalidad del establecimiento. Sin embargo, tampoco en este caso se observaron diferencias estadísticamente significativas entre los grupos definidos. Para Física, el test de *Kruskal–Wallis* arrojó $\chi^2 = 3,72$ ($df = 5$; $p = 0,5905$), mientras que para Química se obtuvo $\chi^2 = 2,16$ ($df = 5$; $p = 0,8259$). Estos resultados son coherentes con los análisis anteriores y permiten afirmar que, en estas cohortes, la combinación de dependencia y modalidad, no explica variaciones relevantes en las notas finales de FIS y QUI.

Finalmente, se evaluó la influencia de la *cohorte* (AÑO) sobre el rendimiento en ambas asignaturas. En el caso de Química, el test de *Kruskal–Wallis* no mostró diferencias significativas entre años ($\chi^2 = 5,14$; $df = 4$; $p = 0,273$), lo que sugiere una relativa estabilidad en las medianas a lo largo de las cohortes analizadas. En contraste, para Física se observaron diferencias altamente significativas entre

cohortes ($\chi^2 = 60,53$; $df = 4$; $p = 2,245 \times 10^{-12}$), lo que indica una fuerte heterogeneidad en las notas de FIS según año de ingreso.

El análisis de las pruebas no paramétricas muestra que ni la dependencia, ni la modalidad del establecimiento escolar, ni su interacción explican diferencias relevantes en las notas de Física y Química en esta muestra. En cambio, las cohortes de ingreso sí difieren significativamente en Física, pero no en Química, lo que posiciona al año de ingreso como un factor contextual clave a considerar en el modelado predictivo de FIS, mientras que el rendimiento en QUI parece ser más estable entre cohortes.

4.4 Modelado predictivo

Los resultados obtenidos en las etapas previas proporcionan un contexto claro para la definición de la estrategia del modelamiento predictivo. Esto, debido a la ausencia de normalidad en la mayoría de las variables predictoras y la presencia de *outliers*, será necesario utilizar métodos no paramétricos y robustos, tales como *Random Forest*, *Gradient Boosting*, *Support Vector Regression* y *k-NN*. No obstante, es necesario evaluar la colinealidad entre los predictores mediante *Variance inflation Factor* (VIF), suprimiendo o combinando aquellos con correlaciones elevadas. La validación cruzada *k-fold* permitirá seleccionar el modelo con mejor capacidad predictiva para las calificaciones de las asignaturas de Física y Química, balanceando estabilidad y evitando sobreajustes.

4.4.1 Análisis de correlación no paramétricas de variables predictoras

Debido a que la mayoría de las variables cuantitativas no presentan una distribución normal y que además se comprobó la presencia de valores atípicos, se aplicó el test de correlación de *Spearman* (ρ) y *Kendall* (τ), los cuales permiten evaluar asociaciones monótonas sin asumir normalidad y son más robustos frente a la presencia de valores *outliers*. Los resultados de los test se muestran en la **tabla 5**.

Tabla 5. Coeficientes de correlación de Spearman y Kendall para la variable objetivos FIS y QUI.

Variable	FIS		QUI	
	<i>Spearman</i>	<i>Kendall</i>	<i>Spearman</i>	<i>Kendall</i>
PTJE LENG	0.080	0.053	0.250	0.177
PTJE MAT	0.344	0.239	0.383	0.272
PTJE MAT2	0.064	0.055	0.301	0.208
PTJE CS	0.246	0.165	0.192	0.135
PD MAT	0.133	0.091	0.196	0.141
PD LEC	0.026	0.016	0.216	0.149
PD FIS	-0.102	-0.075	0.113	0.081
PD QUI	0.161	0.113	0.188	0.134
PD BIO	0.133	0.090	0.251	0.176
FIS	-	-	0.460	0.336

Los resultados muestran que el rendimiento de Física presenta asociaciones positivas de distinta magnitud con un conjunto acotado de variables académicas, particularmente aquellas vinculadas a áreas científicas y matemáticas. Entre las pruebas de ingreso, PTJE_CS y PTJE_MAT presentan las correlaciones más elevadas con FIS. En contraste, PTJE LENG y PTJE_MAT2 muestran asociaciones de menor magnitud.

En relación con las pruebas diagnósticas, se observa que PD_FIS presenta una de las correlaciones más altas con la asignatura de Física. Asimismo, PD_MAT presenta una asociación positiva relevante. Las variables PD_LEC, PD QUI y PD_BIO exhiben correlaciones positivas de menor magnitud en comparación con las anteriores.

El análisis de correlaciones para la asignatura de Química revela un patrón relacional de mayor número y magnitud de asociaciones en comparación con Física. Entre los puntajes de ingreso, PTJE LENG presenta una asociación consistente con QUI, registrando una de las correlaciones más altas dentro de este conjunto de variables. PTJE_CS y PTJE_MAT también presentan correlaciones positivas, aunque de menor magnitud relativa. En el bloque de evaluaciones de diagnósticos, PD_MAT, PD_FIS y PD_BIO evidencian asociaciones positivas y estadísticamente relevantes con QUI. En comparación PD QUI muestra una correlación positiva de menor magnitud respecto del resto de las evaluaciones diagnósticas consideradas.

Se observa una correlación positiva entre FIS y QUI, la cual aparece de manera consistente en ambos coeficientes analizados. Este hallazgo sugiere una coexistencia académica entre ambas asignaturas dentro del conjunto de datos analizados.

4.4.2. Análisis de colinealidad de variables predictoras

Para evaluar la colinealidad de las variables predictoras, se construyó una tabla con los factores de inflación de la varianza (VIF) y cuyos resultados se muestran en la **tabla 6**. Los valores VIF de todos los predictores son mayores a 15 y al menos 5 de ellos son mayores a 30. Estos valores indican que estos predictores tienen redundancia fuerte con otros predictores, esto implica que puede inducir inestabilidad en modelos paramétricos como regresión lineal. además, una alta colinealidad sugiere que estas variables capturan constructos similares, como por ejemplo PD_MAT y PTJE_MAT, capturan información relacionada con habilidades matemáticas previas; PD_LEC y PTJE LENG, hacen lo mismo con habilidades lingüísticas y finalmente PTJE_CS, PD_FIS, PD_BIO, PD_QUI, se relacionan con una dimensión científica común en los estudiantes. Por lo tanto, bajo estas condiciones los valores VIF respaldan la decisión de emplear modelos no paramétricos o modelos robustos hacia la colinealidad y presencia de *outliers*.

Tabla 6. Valores de Variation inflation factor (VIF) VIF de las variables predictoras.

Variable	VIF
PD_LEC	50.93
PTJE_CS	50.20
PD_MAT	43.43
PTJE_MAT	43.31
PD_QUI	31.28
PD_FIS	17.70
PTJE LENG	16.68
PTJE_MAT2	15.97
PD_BIO	15.74

4.4.3 Entrenamiento de algoritmos predictivos

A partir de los resultados de los análisis previos, que muestran variables predictoras que no cumplen con el supuesto de normalidad, la presencia de *outliers* y la fuerte colinealidad, se concluye que la estrategia más apropiada, es utilizar modelos no paramétricos robustos como:

- Random Forest (RF)
- Support Vector Regression (SVM)
- XGBoost (XGB)
- Regresión Ridge
- Regresión Lasso

Por otra parte, para cada variable objetivo se evaluaron escenarios diferenciados los que incluyen, puntajes pruebas estandarizadas (PTJE), notas pruebas diagnósticas (PD) y (MIXTO: PTJE + PD). De esta forma, el conjunto de variables predictoras para FIS y QUI, quedó de la siguiente forma

- **FIS:** PTJE, PD y MIXTO
- **QUI:** PTJE + FIS, PD + FIS y MIXTO + FIS

4.4.4 Resultados entrenamiento para FIS y QUI

Los resultados obtenidos para FIS indican que los modelos presentan una capacidad predictiva limitada en el conjunto de test, con valores de R^2 cercanos a cero en todos los escenarios evaluados, como se muestra en la **tabla 7**.

En términos de validación cruzada, los modelos muestran desempeños moderados ($CV_RMSE \approx 0.11-0.12$), especialmente en los escenarios PTJE y MIXTO, lo que sugiere que existe señal predictiva en los datos de entrenamiento. Sin embargo, esta señal no se generaliza adecuadamente al conjunto de test, evidenciado por el fuerte deterioro del R^2 fuera de muestra. Por otro lado, el escenario PD muestra el desempeño más bajo de manera consistente, lo que concuerda con los análisis de correlación previos, donde las variables diagnósticas presentaron asociaciones débiles con FIS. La incorporación de variables mixtas como PTJE + PD, no produjo mejoras sustantivas respecto al uso exclusivo de PTJE, lo que sugiere una redundancia informativa entre ambos bloques de variables.

Estos resultados muestran que la capacidad explicativa de los puntajes estandarizados y de los diagnósticos institucionales respecto del rendimiento en Física es limitada y variable entre los modelos evaluados, sin evidenciar un patrón consistente de asociación.

Tabla 7. Métricas evaluación modelos predictivos para la variable FIS.

Escenario	Modelo	CV_RMSE	CV_MAE	CV_R ²	Test_RMSE	Test_MAE	Test_R ²
PTJE	RF	0.113	0.0905	0.2203	0.1434	0.1177	0.0031
PTJE	SVM	0.115	0.0911	0.1799	0.1499	0.1233	0.0006
PTJE	RIDGE	0.113	0.0887	0.1963	0.1467	0.1207	0.0018
PTJE	LASSO	0.112	0.0886	0.2178	0.1431	0.1199	0.0044
PTJE	XGB	0.121	0.0933	0.2301	0.1519	0.1293	0.0109
PD	RF	0.124	0.1018	0.0435	0.1355	0.1143	0.0017
PD	SVM	0.124	0.1022	0.1057	0.1427	0.1203	0.0015
PD	RIDGE	0.125	0.1026	0.0595	0.1363	0.1165	0.0063
PD	LASSO	0.124	0.1028	0.0407	0.1327	0.1140	0.0001
PD	XGB	0.143	0.1161	0.0504	0.1549	0.1283	0.0014
MIXTO	RF	0.112	0.0885	0.2239	0.1374	0.1154	0.0059
MIXTO	SVM	0.114	0.0892	0.2177	0.1489	0.1256	0.0003
MIXTO	RIDGE	0.113	0.0875	0.2187	0.1476	0.1238	0.0001
MIXTO	LASSO	0.111	0.0869	0.2420	0.1430	0.1204	0.0031
MIXTO	XGB	0.125	0.0975	0.1838	0.1486	0.1236	0.0114

En contraste, los modelos utilizados para QUI exhiben un comportamiento sustancialmente diferente, ver **tabla 8**. En todos los escenarios evaluados, los modelos alcanzan valores de R² positivos y estables tanto en validación cruzada como en el conjunto utilizado para test, destacando particularmente los escenarios que incluyen FIS como predictor. El mejor desempeño se observa en el escenario MIXTO + FIS, donde los modelos *Lasso*, *Ridge* y *Random Forest* alcanzan valores de R² en test del orden de 0.23–0.26, junto con errores absolutos medios bajos (MAE ≈ 0.04–0.05 en escala normalizada).

Tabla 8. Métricas evaluación modelos predictivos para la variable QUI.

Escenario	Modelo	CV_RMSE	CV_MAE	CV_R ²	Test_RMSE	Test_MAE	Test_R ²
PTJE+FIS	RF	0.0616	0.0492	0.2861	0.0636	0.0476	0.1290
PTJE+FIS	SVM	0.0623	0.0504	0.3124	0.0671	0.0473	0.0859
PTJE+FIS	RIDGE	0.0629	0.0502	0.3183	0.0619	0.0437	0.1939
PTJE+FIS	LASSO	0.0621	0.0485	0.3123	0.0602	0.0425	0.2266
PTJE+FIS	XGB	0.0643	0.0528	0.2859	0.0687	0.0524	0.1200
PD+FIS	RF	0.0664	0.0531	0.1658	0.0602	0.0438	0.2000
PD+FIS	SVM	0.0659	0.0542	0.2533	0.0704	0.0505	0.0156
PD+FIS	RIDGE	0.0648	0.0524	0.2639	0.0620	0.0451	0.1698
PD+FIS	LASSO	0.0629	0.0500	0.2823	0.0611	0.0430	0.1744
PD+FIS	XGB	0.0788	0.0646	0.1038	0.0666	0.0520	0.1354
MIXTO+FIS	RF	0.0640	0.0509	0.2189	0.0584	0.0412	0.2597
MIXTO+FIS	SVM	0.0665	0.0553	0.2139	0.0665	0.0482	0.0545
MIXTO+FIS	RIDGE	0.0633	0.0510	0.2964	0.0600	0.0416	0.2356
MIXTO+FIS	LASSO	0.0622	0.0493	0.3059	0.0596	0.0418	0.2412
MIXTO+FIS	XGB	0.0750	0.0615	0.1277	0.0605	0.0427	0.2125

Estos resultados son coherentes con los análisis de correlación, que mostraron una asociación moderada entre FIS y QUI, lo que respalda empíricamente la inclusión de FIS como predictor en los modelos de Química. Asimismo, la superioridad del escenario mixto sugiere que el desempeño en Química se asocia tanto a competencias académicas generales como a diagnósticos disciplinares, siempre que se controle la colinealidad mediante regularización o técnicas de reducción de dimensionalidad.

Dado que los análisis previos evidenciaron una alta colinealidad interna entre las variables predictoras pertenecientes a los bloques de puntajes estandarizados (PTJE) y pruebas diagnósticas (PD), y considerando además que los modelos basados en la combinación completa de predictores

(MIXTO+FIS) mostraron mejoras marginales pero consistentes en el desempeño predictivo de la variable QUI, se procedió a evaluar técnicas de reducción de dimensionalidad supervisadas y no supervisadas.

En este contexto, se aplicaron los métodos de Regresión por Componentes Principales (PCR) y Regresión por Mínimos Cuadrados Parciales (PLSR), utilizando como conjunto de predictores las variables normalizadas PTJE_N, PD_N y FIS_N, y considerando exclusivamente observaciones con valores válidos en la variable objetivo QUI_N.

El modelo PCR alcanzó su mejor desempeño utilizando 12 componentes principales, seleccionados mediante validación cruzada interna minimizando el error de predicción. En el conjunto de test, este modelo presentó un RMSE de 0.0687, un MAE de 0.0574 y un R^2 de 0.169. Estos resultados indican que, si bien la reducción de dimensionalidad permite mitigar la colinealidad y estabilizar el modelo, una parte relevante de la varianza capturada por los primeros componentes principales no está necesariamente alineada con la variabilidad explicativa de la variable QUI, lo cual es consistente con la naturaleza no supervisada del PCA subyacente al método PCR. Por su parte, el modelo PLSR mostró un desempeño ligeramente superior, alcanzando su mejor ajuste con 2 componentes latentes, también seleccionados mediante validación cruzada. En el conjunto de test, el modelo PLSR obtuvo un RMSE de 0.0682, un MAE de 0.0576 y un R^2 de 0.182.

La mejora observada respecto a PCR, aunque moderada, se evidencia en las métricas de desempeño del modelo, ya que el enfoque PLSR construye componentes que maximizan simultáneamente la covarianza entre los predictores y la variable respuesta. En este contexto, los resultados muestran que una estructura latente de baja dimensionalidad, compartida entre los puntajes estandarizados, las pruebas diagnósticas y el rendimiento en Física, está presente en el modelo aplicado a Química.

La comparación entre PCR y PLSR evidencia diferencias en su desempeño en un contexto de alta colinealidad entre predictores. En particular, PLSR presenta valores levemente superiores a PCR según las métricas evaluadas. Sin embargo, los valores de R^2 indican que una proporción relevante de la variabilidad en QUI no es explicada por ninguno de los modelos, aun considerando técnicas de reducción de dimensionalidad. Estos resultados permiten describir el comportamiento comparativo de ambos métodos en la predicción del rendimiento en Química.

Capítulo V: Discusión.

Los resultados de este estudio permiten comprender con mayor profundidad los factores académicos asociados al rendimiento académico en las asignaturas de Física y Química en estudiantes de la carrera de Educación General Básica de la Universidad de Concepción, Campus Los Ángeles. En primer lugar, los análisis descriptivos evidenciaron brechas importantes en los puntajes de las pruebas diagnósticas institucionales de Física y Química, lo cual indica un dominio insuficiente de conocimientos previos. Estas características son coherentes con lo planteado por Barahona Ibarra (2021), quien afirma que tanto los estudiantes que poseen una formación disciplinar sólida como los que carecen de ella, presentan un bajo dominio conceptual y que la relación entre competencia curricular y rendimiento académico resulta débil. Tal fenómeno evidencia que los estudiantes ingresan con una base disciplinar limitada, lo que repercute en un rendimiento universitario inferior y con mayor variabilidad.

De este modo, se menciona la relevancia de la formación preuniversitaria del estudiante, sin embargo, se constató la ausencia de una diferencia significativa respecto de la dependencia o modalidad del establecimiento de origen de cada estudiante, lo que indica que, una vez en la universidad, el contexto escolar de origen pierde relevancia frente a las competencias académicas personales. Este hallazgo es similar a lo planteado por Llanos Salazar *et al.* (2018), ya que, el rendimiento académico en la educación superior depende principalmente de los factores académicos individuales de cada uno de los estudiantes al momento de ingresar a la educación superior.

Por otro lado, la evaluación mostró anomalías en los supuestos estadísticos y presencia de colinealidad entre los predictores (especialmente habilidades matemáticas y lenguaje), impidiendo la aplicación de modelos lineales tradicionales. Esta limitación metodológica ya había sido identificada por Aguilar Reyes *et al.* (2025), quienes señalaron que los datos académicos reales rara vez cumplen con los supuestos necesarios para aplicar regresión lineal, debiendo optar por modelos logísticos o análisis ROC para clasificar el rendimiento estudiantil. La coincidencia con el presente estudio refuerza la complejidad inherente a datos educativos y la necesidad de emplear enfoques predictivos alternativos, lo cual justifica plenamente la selección metodológica utilizada en esta investigación.

En este contexto, las diferencias observadas entre cohortes, particularmente más marcadas en la asignatura de Física que en Química, permiten plantear la posibilidad de que existan factores específicos

asociados a cada generación de ingreso que no fueron directamente considerados en el presente estudio. Entre estos factores se podrían incluir eventuales cambios curriculares, variaciones en las condiciones de enseñanza, ajustes en los programas de estudio o características propias de los grupos de estudiantes admitidos en cada año académico. Si bien estas variables no fueron medidas de manera explícita, su mención resulta pertinente para interpretar la mayor variabilidad entre cohortes observada en Física, en contraste con el comportamiento más estable evidenciado en Química. Desde una perspectiva interpretativa, esto sugiere que el rendimiento en Física podría ser más sensible a condiciones contextuales y estructurales del proceso formativo, mientras que Química presentaría una mayor homogeneidad en sus resultados, independientemente del año de ingreso.

En las asignaturas de Física y Química, la caracterización de los factores académicos de ingreso permitió identificar que las variables asociadas a las habilidades matemáticas y científicas son las que tienen mayor capacidad predictiva sobre el rendimiento académico, especialmente en la asignatura de Física. Además, los análisis acerca de la asignatura de Física, muestran diferencias estadísticamente significativas entre las cohortes, lo que sugiere que en esta disciplina es más sensible al nivel de preparación previa de cada estudiante, sin embargo, la asignatura de Química muestra resultados más estables, lo que podría atribuirse a un enfoque más conceptual y menos dependiente de las habilidades matemáticas específicas. Tal como lo plantea Jiménez Figueroa (2025), la asignatura de Física requiere un dominio lógico-matemático más desarrollado, el cual es difícil de compensar una vez iniciado el proceso universitario.

Por último, las diferencias entre cohortes y el comportamiento desigual de los predictores confirman que el rendimiento académico es un fenómeno multifactorial, influido tanto por las capacidades previas de los estudiantes como por la naturaleza disciplinar y las estrategias pedagógicas empleadas. Este resultado es coherente con lo señalado por Fernández Zalazar et al. (2024), quienes destacan la necesidad de adaptar las estrategias educativas a las particularidades de cada asignatura y a las características de los estudiantes. En este contexto, los hallazgos de la presente investigación pueden contribuir al diseño de sistemas de alerta temprana y al desarrollo de estrategias pedagógicas diferenciadas, especialmente en aquellas asignaturas con menor capacidad predictiva a partir de indicadores académicos previos.

Capítulo VI: Conclusiones.

La pregunta de investigación planteada en este estudio buscó determinar en qué medida los antecedentes académicos de ingreso y las evaluaciones diagnósticas institucionales permiten predecir el rendimiento académico en las asignaturas de Física y Química en estudiantes de primer año de Educación General Básica.

El análisis exploratorio y los test de correlación no paramétricos evidenciaron una alta colinealidad interna dentro de los grupos de variables estandarizadas de ingreso (PTJE) y de las pruebas diagnósticas (PD), así como una marcada no normalidad y presencia de outliers reales. Las correlaciones de Spearman y Kendall muestran asociaciones bajas a moderadas con el rendimiento en las asignaturas de Física y Química, siendo más consistentes en esta última asignatura, especialmente para los puntajes de Matemática y Lenguaje, y para el rendimiento previo en Física.

A partir de los resultados obtenidos, se concluye que la capacidad predictiva de dichas variables es parcial y diferenciada según la asignatura. Mientras que en Química se observa una relación consistente, aunque moderada, entre los antecedentes académicos y el rendimiento posterior, en la asignatura de Física esta relación resulta débil y altamente dependiente del contexto, sin una capacidad de generalización adecuada en los conjuntos de test. En consecuencia, la pregunta de investigación es respondida afirmativamente solo de manera parcial, evidenciando que los factores académicos de ingreso no explican de forma íntegra el desempeño en ambas asignaturas.

La comparación entre escenarios de predicción basados únicamente en puntajes estandarizados, únicamente en pruebas diagnósticas y en la combinación de ambos conjuntos permitió establecer que no existe un conjunto universalmente óptimo para ambas asignaturas. En la asignatura de Física, ninguno de los escenarios evaluados logra un desempeño predictivo satisfactorio en test, mientras que en la asignatura de Química los escenarios mixtos que incorporan el rendimiento de la asignatura de Física presentan los mejores resultados. Este hallazgo respalda la hipótesis de que el rendimiento en Química depende en mayor medida de competencias científicas previas consolidadas durante el curso de Física.

En el escenario mixto para Química, la aplicación de técnicas de reducción de dimensionalidad como PCR y PLSR permitió estabilizar los modelos frente a la colinealidad existente entre predictores.

En particular, PLSR mostró un desempeño ligeramente superior, al integrar de forma supervisada la información de las variables predictoras. Sin embargo, estas técnicas no incrementaron de manera sustantiva la varianza explicada, lo que refuerza la conclusión de que la capacidad predictiva del modelo está limitada por la naturaleza multifactorial del rendimiento académico.

Finalmente, los resultados de este estudio evidencian que los antecedentes académicos de ingreso constituyen un insumo relevante pero insuficiente para explicar el rendimiento en asignaturas científicas de primer año, especialmente en Física. La evidencia empírica respalda la necesidad de enfoques diferenciados por asignatura y sugiere que futuras investigaciones deben incorporar variables pedagógicas, curriculares y contextuales para lograr modelos predictivos más robustos y explicativamente coherentes.

Limitaciones y proyecciones del estudio.

En primera instancia, una de las limitaciones fundamentales de este estudio radicó en el proceso de recolección de los insumos necesarios para la construcción de la base de datos. Esta etapa presentó desafíos significativos en la consolidación de la información académica, elemento crítico sobre el cual se ejecutó el análisis estadístico y la posterior evaluación del modelo de predicción propuesto. La complejidad en la obtención de registros históricos completos y centralizados supuso una barrera inicial que requirió de esfuerzos adicionales para garantizar la integridad y robustez de los datos analizados.

Por otro lado, el alcance predictivo de los modelos desarrollados en este estudio se encuentra condicionado, en primer lugar, por la naturaleza multifactorial del rendimiento académico en ciencias, el cual depende de dimensiones cognitivas, pedagógicas y contextuales que no pueden ser capturadas completamente mediante indicadores cuantitativos estandarizados. En consecuencia, una fracción significativa de la variabilidad observada en los resultados de Física y Química permanece asociada a factores que no fueron considerados en este estudio, lo que limita el valor máximo alcanzable de las métricas explicativas, incluso al emplear algoritmos de aprendizaje automático más robustos.

Como proyección metodológica, futuras investigaciones podrían beneficiarse de la incorporación de variables derivadas con mayor significado pedagógico, del uso de esquemas de validación cruzada repetida para mejorar la estabilidad de las estimaciones y de la integración de técnicas avanzadas de reducción de dimensionalidad o aprendizaje conjunto cuando se disponga de conjuntos de datos más amplios. Adicionalmente, la inclusión de información contextual o cualitativa permitiría avanzar hacia modelos explicativos más integrales, complementando el enfoque predictivo desarrollado en este estudio.

Referencias.

- Acosta Faneite, S. F. (2023). Los paradigmas de investigación en las Ciencias Sociales: Capítulo 4. *Editorial Idicap Pacífico*, 60-79.
- Agila Cuenca, M. A. (2020). *Factores educativos asociados al rendimiento académico en el área de Química de la Carrera de Pedagogía de las Ciencias Experimentales, Química y Biología de la Facultad de Filosofía, Letras y Ciencias de la Educación de la Universidad Central del Ecuador, 2019* [Tesis de licenciatura, Universidad Central del Ecuador]. Repositorio Institucional.
- Aguilar-Reyes, J. E., Mejía-Peñañiel, E. F., Morocho-Barrionuevo, T. P., & Velasco Castelo, G. M. (2025). Estudio del rendimiento académico mediante la comparación de modelos de regresión y árboles de clasificación. *Telos*, 27(1), 94-115.
- Ávila-Morales, J., Becerril-Santos, A., Vega-Malagón, A., Vega-Malagón, G., Camacho-Calderón, N., & Leo-Amador, G. E. (2014). Paradigma 1 en la investigación: Enfoque cuantitativo y cualitativo. *European Scientific Journal*, 10(15), 524–539.
- Barahona Ibarra, A. E. (2021). *Evaluación de la competencia curricular y el rendimiento académico de la asignatura de química de estudiantes universitarios y de bachillerato* [Tesis doctoral, Universitat d'Alacant/Universidad de Alicante].
- Barahona Ibarra, A. E., Navas Martínez, L., Veas Iniesta, A., & Pozo Zapata, R. F. (2024). La competencia curricular y el rendimiento académico de química de estudiantes universitarios y de bachillerato. *Ciencia Latina Revista Científica Multidisciplinar*, 8(4).
- Barahona U., P. (2014). Factores determinantes del rendimiento académico de los estudiantes de la Universidad de Atacama. *Estudios Pedagógicos (Valdivia)*, 40(1), 25–39.
- Barrera Hernández, A. E., Garza Hernández, L. G., & Ramos Cárdenas, S. (2019). Bienestar psicológico y rendimiento académico en estudiantes universitarios. *Revista Electrónica de Psicología Iztacala*, 22(2), 594–617.
- Barraza, R., Muñoz-Navarro, R., & Behrens, M. A. (2022). Bienestar psicológico y rendimiento académico en estudiantes universitarios: Una revisión sistemática. *Ciencia y Técnica*, 27(1), 101-112.
- Bastías-Bastías, L. S., & Iturra-Herrera, C. (2022). La formación inicial docente en Chile: Una revisión bibliográfica sobre su implementación y logros. *Revista Electrónica Educare (Educare Electronic Journal)*, 26(1), 1–22.
- Baudelio, L. L., González-Palacios, A., González-Álvarez, M. de los Á., & Martínez-González, M. G. (2014). Fracaso escolar: Conceptualización y perspectivas de estudio. *Revista de Educación y Desarrollo*, (pág. 81).

- Bello, F. A., Kohler, J., Hinrichsen, K., Araya, V., Hidalgo, L., & Jara, J. L. (2020). Using machine learning methods to identify significant variables for the prediction of first-year Informatics Engineering students dropout. *2020 IEEE International Conference on Big Data*.
- Biggs, J. (2001). El aprendizaje académico visto desde la perspectiva del modelo 3P de John Biggs. *Revista de Investigación Educativa*, 19(1), 23–46.
- Bórquez Mella, J., Sanhueza Vega, T., Alarcón Chávez, P., & Díaz Garrido, P. (2023). Desarrollo de competencias en la formación inicial de docentes: Percepción de las personas tituladas de una universidad chilena. *Actualidades Investigativas en Educación*, 23(2), 1-28.
- Campos Palacios, E. W. (2025). El aprendizaje autorregulado de estudiantes universitarios de un programa de pedagogía chileno. *Revista Boletín Redipe*, 14(8), 94-105.
- Cañizales Mendoza, M. Á., González Murillo, M., & Foster Vega, M. (2025). *Habilidades cognitivas en el rendimiento académico en matemática para estudiantes universitarios de primer ingreso*. REDEPSIC, 4(2), 78–93.
- Castrillón, O., Sarache, W., & Ruiz-Herrera, S. (2020). Predicción del rendimiento académico por medio de técnicas de inteligencia artificial. *Formación Universitaria*, 13(1), 93-102.
- Ceja Oseguera, S., Cervantes Contreras, N., & Ramírez Murillo, L. M. (2011). Estudio de la violencia que el maestro de educación media superior ejerce sobre sus alumnos como factor de desmotivación académica. *Métodos. Revista Electrónica de Investigación Aplicada en Derechos Humanos*, 1, 46-65. Comisión de Derechos Humanos de la Ciudad de México.
- Cervini, R. (2018). Factores personales-institucionales que impactan el rendimiento académico. *Revista de Investigación Educativa*, 36(2), 61-86.
- Covarrubias Apablaza, C. G., Veas Alfaro, M., & González Marino, I. (2023). Autorregulación del aprendizaje en estudiantes de derecho de primer año: Un diseño preexperimental. *Revista Pedagogía Universitaria y Didáctica del Derecho*, 10(1), 195–214.
- Cobo-Rendón, R., López-Angulo, Y., Sáez-Delgado, F., & Mella-Norambuena, J. (2022). Engagement, motivación académica y ajuste de estudiantado universitario. *Revista Electrónica Educare (Educare Electronic Journal)*, 26(3), 1-19.
- Colmenares, M., & Delgado, F. (2008). Aproximación teórica al estado de la relación entre rendimiento académico y motivación de logro en educación superior. *Revista de Ciencias Sociales*, 14(3), 306–319.
- Consejo de Rectores de las Universidades Chilenas (CRUCH). (2024). *ENCE 2023: Compromiso estudiantil y uso del tiempo impactan directamente en el rendimiento académico*.

- Contreras, F., Espinosa, J. C., Esguerra, G. A., Haikal, A., Ramírez-Salazar, A. F., & UribeRodríguez, A. F. (2021). Resiliencia y estrategias de afrontamiento en estudiantes universitarios. *Psicogente*, 24(46), 1-19.
- Del Carpio-Mendoza, R. A. (2024). Predicción del rendimiento académico utilizando modelos de aprendizaje automático: Una revisión sistemática de la literatura. *593 Digital Publisher CEIT*, 9(6), 1038–1054.
- De Luca, J. M. N., Robles, J. M. G., Ibujés, M. O. S., & Tasigchana, M. A. S. (2025). Optimización del aprendizaje conceptual y práctico en matemáticas, física y química mediante la implementación de tecnologías digitales y estrategias de gamificación en la educación superior. *Revista Social Fronteriza*, 5(3), 4-5.
- Díaz Herrera, C., López Espinoza, M. Á., & Pino Villalón, J. L. (2020). *Midiendo capital cultural interiorizado: una escala para educación superior* (pp. 4–6). *Revista Pedagogía Universitaria y Didáctica del Derecho*, 7(2), 302-304.
- Díaz-Landa, B., Meleán-Romero, R., & Marín-Rodríguez, W. (2021). Rendimiento académico de estudiantes en educación superior: Predicciones de factores influyentes a partir de árboles de decisión. *TELOS: Revista de Estudios Interdisciplinarios en Ciencias Sociales*, 23(3), 617620.
- Duche, J., Gallardo, M., & González, A. (2023). *Transición secundaria-educación superior: Desafíos para los estudiantes, desafíos para las instituciones*. *Revista de Educación Superior*, 52(1), 45–62.
- Durán, J., & Crispín, D. (2020). Influencia de la autoestima en el rendimiento académico de estudiantes de la carrera de Medicina de la Universidad Mayor de San Andrés. *Revista Médica La Paz*, 26(2), 9–15.
- Escuela de Profesores del Perú. (2024). *Logro de aprendizaje, concepto, aspectos generales, mejoras, retos y perspectivas*.
- Espinoza, O. (2017). Rendimiento académico, factores socioeconómicos y la educación superior en Chile. *Revista Iberoamericana de Educación Superior*, 8(21), 108–122.
- Fernández Casal, R., Costa Bouzas, J., & Oviedo de la Fuente, M. (2024). *Métodos predictivos de aprendizaje estadístico*, 97.
- Fernández Zalazar, D. C., Matos, F. B., & Frangullo, Y. B. (2024). Analíticas de aprendizaje, herramientas y estrategias de enseñanza [Presentación de un trabajo/póster]. XVI Congreso Internacional de Investigación y Práctica Profesional en Psicología. XXXI Jornadas de Investigación. XX Encuentro de Investigadores en Psicología del MERCOSUR. VI Encuentro de Investigación de Terapia Ocupacional. VI Encuentro de Musicoterapia, Buenos Aires, Argentina.

- Garbanzo Vargas, G. M. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Revista Educación*, 31(1), 47-48
- Garcés Bustamante, J., Labra Godoy, P., & Vega Guerrero, L. (2020). La retroalimentación: una estrategia reflexiva sobre el proceso de aprendizaje en carreras renovadas de educación superior. *Cuadernos de Investigación Educativa*, 11(1).
- García Arista, A., Torres-Segoviano, C. A., & Velázquez Olmedo, C. E. (2024). Rendimiento académico y percepción del clima escolar en estudiantes de bachillerato. *Revista de Psicología de la Universidad Autónoma del Estado de México*, 13(37), 223–238.
- Garzon, R., Rojas, M. O., Riesgo, L. D., Pinzón, M., & Salamanca, A. L. (2010). Factores que pueden influir en el rendimiento académico de estudiantes de Bioquímica que ingresan en el programa de Medicina de la Universidad del Rosario-Colombia. *Educación médica*, 13(2), 85-96.
- González, A., & Muñoz, R. G. (2021). Capital cultural y rendimiento académico de estudiantes normalistas en Sonora, México. *CPU-e, Revista de Investigación Educativa*, (32), 53-73.
- González, P., & Muñoz, A. (2019). Factores asociados al fracaso académico en estudiantes de primer año. *Revista de Educación Superior*, 48, 51.
- González Sanzana, A., & Arce Secul, R. (2021). Factores personales y de acceso que inciden sobre la permanencia y deserción universitaria en estudiantes de pedagogía en una universidad chilena de zona geográfica extrema. *Sophia Austral*, 27(1), 1-18.
- Guillart, M. M. R., López, M. M. C., & del Pino Caro, L. M. (2020). La gestión del conocimiento en la enseñanza de la Física para la formación de profesionales. *Latin-American Journal of Physics Education*, 14(4), 6.
- Henríquez Cabezas, N. & Vargas Escobar, D. (2022). Modelos predictivos de rendimiento y deserción académica en estudiantes de primer año de una universidad pública chilena. *REXE: Revista de Estudios y Experiencias en Educación*, 21(45), 312-313.
- Hernández Ávila, C. E., & Carpio, N. (2019). Introducción a los tipos de muestreo. *Alerta*, 2(1), 7579.
- Hernández Knipp, C., & Zúñiga Lalanne, D. (2019). *Estudiantes universitarios de primera generación en Chile: una revisión bibliográfica* (Seminario de título, Universidad de Valparaíso, Facultad de Medicina, Escuela de Psicología). Valparaíso, Chile. 34-36.
- Hernández-Sampieri, R., Fernández-Collado, C., & Baptista-Lucio, M. del P. (2014). *Metodología de la investigación* (6ta ed.). McGraw-Hill Education. 122-127.

- Hernández-Yépez, P. J., Contreras-Carmona, P. J., Inga-Berrospi, F., Basurto Ayala, P., & Valladares-Garrido, M. J. (2022). Factores asociados al rendimiento académico en estudiantes de medicina. *Revista Cubana de Medicina Militar*, 51(1).
- Herrera Rivera, P., & Arancibia Carvajal, S. (2022). Modelo exploratorio de factores que inciden en el rendimiento académico percibido. *Revista de Estudios y Experiencias en Educación (REXE)*, 21(45), 346-347.
- Hurtado-Palomino, A., Merma-Valverde, W., Ccorisapra-Quintana, F. de M., Lazo-Cerón, Y., & Boza-Salas, K. (2021). Estrategias de enseñanza docente en la satisfacción académica de los estudiantes universitarios. *Comuni@ccion: Revista de Investigación en Comunicación y Desarrollo*, 12(3), 217-228.
- Jérez Yáñez, O., Aranda Contreras, C., Hasbún Held, B., Lizama Bucarey, C., Marilao Carrasco, C., Rittershausen Klaunig, S., & Rojas Pino, M. (2018). *Potenciando mi aprendizaje en el primer año de universidad: Guía para desarrollar estrategias de estudio* (pp. 7–8, 31, 48– 52). Universidad de Chile.
- Jiménez Figueroa, Á. S. (2025). Aprendizaje Basado en la Resolución y Explicación de Problemas (ABREP) para estudiantes de cursos de Física en la Universidad Católica de Temuco, Chile. *Revista Chilena De Educación Científica*, 26(1), 100–111.
- Kuzmanic Vidal, D. M.; Valenzuela, J. P.; Villalobos, C.; Quaresma, M. L. (2021). *Socioeconomic Segregation in Higher Education: Evidence for Chile (2009–2017)*. *Higher Education Policy*.
- Landin Garcia, M. E., Lima Balcázar, P. A., & Mena Gallardo, A. A. (2024). Aprendizaje Basado en Problemas en el Desarrollo de Habilidades de Pensamiento Crítico de los Estudiantes del Séptimo Ciclo de la Carrera de Pedagogía de las Ciencias Experimentales Química y Biología. *Ciencia Latina Revista Científica Multidisciplinar*, 8(2).
- Lara-García, B., González-Palacios, A., González-Álvarez, M. de los Á., & Martínez-González, M. G. (2014). Fracaso escolar: Conceptualización y perspectivas de estudio. *Revista de Educación y Desarrollo*, 30, 81.
- Lerner Matiz, J. (2010). Rendimiento académico en la perspectiva del desarrollo y bienestar del estudiante: Una visión clínica psicoanalítica. *Revista Universidad EAFIT*, 46(160), 73-89.
- Llanos Salazar, E. C., Medina Boguer, V. A., Muñoz Ortiz, M. P., Karmach Sánchez, E., & Herrera Cofré, R. (2018). Análisis de los perfiles de ingreso y su relación con el rendimiento académico. *Multidisciplinary Business Review*, 11(1), 13–26.
- López Aguado, M., & López Alonso, A. I. (2013). Los enfoques de aprendizaje. Revisión conceptual y de investigación: Theoretical and Research Review. *Revista colombiana de educación*, (64), 136.

- López-Angulo, Y., Pérez-Villalobos, M. V., Cobo-Rendón, R. C., & Díaz-Mujica, A. E. (2020). Apoyo social, sexo y área del conocimiento en el rendimiento académico autopercebido de estudiantes universitarios chilenos. *Formación Universitaria*, 13(3), 11–20.
- Martínez-Corona, J. I., Palacios-Almón, G. E., & Oliva-Garza, D. B. (2023). Guía para la revisión y el análisis documental: Propuesta desde el enfoque investigativo. *Ra Ximhai*, 19(1), 67–83.
- Martínez-Otero Pérez, V. (2021). *Rendimiento escolar y formación integral*. Ediciones Octaedro.
- Mc Coll Calvo, P., Goset Poblete, J., Martínez Lomakin, F., Searle Solar, M., & Silva Orrego, V. (2021). Modelo predictivo de riesgo académico en estudiantes de medicina. *Revista Médica de Chile*, 149(12), 1788–1789.
- Mella Norambuena, J., Badilla-Quintana, M. G., & López Angulo, Y. (2022). Modelos predictivos basados en uso de analíticas de aprendizaje en educación superior: una revisión sistemática. *Texto Livre: Linguagem e Tecnologia*, 15(1), e36310, 18.
- Ministerio de Educación - Centro de Estudios. (2023). *Matrícula oficial 2023: Análisis de base matrícula por estudiante* (Apuntes de Educación N°34), 2.
- Morales Navarro, M. A., Rojas Lillo, L. A., Guzmán Utreras, E. A. E., & Baeza Ugarte, C. G. (2022). Estrés, autoeficacia, apoyo-social y personalidad en estudiantes del Programa de Acceso y Acompañamiento Efectivo a la Educación Superior. *Revista Educación*, 46(1), 1-24.
- Morel, A. (2025, 14 de enero). *SLEP registraron peor promedio en PAES que municipales y subvencionados y establecimientos públicos no superaron los 600 puntos*. *The Clinic*.
- Muñoz Muñoz, E., Sandoval Manríquez, M., & González Bravo, L. (2016). *Explorando asociaciones entre capital cultural y rendimiento académico en alumnos de primer año de Medicina y Tecnología Médica en la Universidad San Sebastián, Concepción*. *TS Cuadernos de Trabajo Social*, (14), 46–78.
- Murillo-García, O. L., & Luna-Serrano, E. (2020). Autorregulación del aprendizaje en estudiantes universitarios inscritos en programas de formación docente. *Revista de Educación y Desarrollo*, (52), 65–75.
- Novoa Mora, F. E. (2024). *Narrativas de persistencia: Factores que influyen en las trayectorias académicas de estudiantes provenientes de contextos vulnerables en la educación superior* (Tesis de maestría, Universidad de Concepción). Universidad de Concepción, Facultad de Ciencias Sociales, 170-173.
- Núñez Villalobos, D. A. (2025). Modelo predictivo basado en aprendizaje automático para la retención estudiantil en educación superior. *European Public & Social Innovation Review*, 10, 1-21.

- Orejarena-Silva, H. A. (2020). Relación entre inteligencia emocional, estilos de aprendizaje y rendimiento académico en un grupo de estudiantes de psicología. *Inclusión & Desarrollo*, 7(2), 26.
- Ortiz Ojeda, A. F., & Canto Herrera, P. J. (2013). Estilos de aprendizaje y rendimiento académico en estudiantes de ingeniería en México. *Revista Estilos de Aprendizaje*, 6(11), 165.
- Peralta de la Vega, M. (2005). La formación académica en secundaria y en el centro preuniversitario y su influencia en el rendimiento académico en el área de ciencias naturales de la UNSCH. Tesis de maestría, Universidad Nacional de San Cristóbal de Huamanga. *Repositorio Institucional*.
- Pérez-Herrera, M. A., Jiménez-Pérez, N. J., & De la Hoz-Vallejo, M. (2022). El capital cultural incorporado y su relación con el rendimiento académico de los estudiantes del programa de sociología de la Universidad del Atlántico. *Revista Encuentros*, 20(01).
- Pérez, M. M. M., Soto-Ardila, R. (2022). Factores que influyen en el rendimiento académico universitario: Una perspectiva integral. *Revista Electrónica Educare*, 26(1), 1-17.
- Quintanal Pérez, F., (2023). Aprendizaje basado en problemas para Física y Química de Bachillerato. Estudio de caso. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 20(2), 220101.
- Reusche, R. (1995). Estructura y funcionamiento familiar en un grupo de estudiantes de secundaria de nivel socioeconómico medio con alto y bajo rendimiento escolar. *Revista de Psicología, UNIFE*, 3, 163-190.
- Román-González, M., Zapata-Ros, M., & Alarcón-Mendoza, A. (2020). Desafíos y predictores del rendimiento académico en el primer año universitario. *Revista Iberoamericana de Educación Superior*, 11(31), 107-124.
- Ruiz, M. V., & Torres-Soto, A. (2020). Calidad de los aprendizajes de estudiantes de pedagogía: Influencia del modelo educativo. *Revista Fuentes*, 22(2), 238-250.
- Salgado Díaz, G. (2022). La naturaleza cultural del contenido biológico y su evaluación integradora en la formación del profesor de Biología. *Revista Didasc@lia: didáctica y educación*, 13(2), 244–256.
- Salinas, V., Zea, K., Salinas, J., & Jácome, T. (2025). Impacto de la autoestima en el desempeño académico en educación superior: Revisión sistemática y meta-análisis de la correlación en estudiantes universitarios. *Revista REG*, 4(3), 274–297.
- Sánchez Jiménez, C. (2024). Uso de la agregación Bootstrap y los bosques aleatorios en la minería de datos: Creación de una aplicación web con R para clasificar o predecir datos reales (Trabajo de fin de grado). *Universidad de Salamanca*.

- Schmidt, B. J., Boero, P., & Méndez Vera, J. A. (2023). Factores que influyen en la deserción universitaria: El caso de una universidad estatal chilena. *Revista Portuguesa de Educación*, 36(1),
- Soza Mora, S. E. (2021). Factores asociados a la calidad del rendimiento académico de estudiantes en la educación superior. *Revista Ciencias de la Salud y Educación Médica*, 3(3), art. 5, 40.
- Tourón, J., & Santiago, R. (2015). El modelo Flipped Learning y el desarrollo de la competencia de aprender a aprender. *Profesorado. Revista de Currículum y Formación de Profesorado*, 19(3), 1-17.
- Universidad de Concepción. (2024). *Modelo educativo de la Universidad de Concepción Ciudad Universitaria*.
- Valle, G., Giménez, M., Suligoy, M., Raiano, J., Gómez, G., García, S., Aciar, M., Bazán, A., & Fierro, S. (2014). Éxito y fracaso académico desde la perspectiva de los estudiantes universitarios. *Dos Puntas*, VI(9), 162-162.
- Vega Vargas, M. F. (2025). *Relación entre el bienestar subjetivo y el apego a la universidad: Impacto en la experiencia estudiantil y rendimiento académico en estudiantes de primer año de Trabajo Social en Concepción, Chile* (tesis de Magíster, Universidad de Concepción), 61.
- Vera Sagredo, A., Cerda Etchepare, G., Aragón Mendizábal, E., y Pérez Wilson, C., (2021). Rendimiento académico y su relación con variables socioemocionales en estudiantes chilenos de contextos vulnerables. *Educación XXI*, 24(2), 375-398.
- Zambrano-Matamala, C., Díaz-Mujica, A., Perez-Villalobos, M. V., & Rojas-Díaz, D. (2020). Análisis de estrategias de autorregulación en estudiantes de pedagogía de una universidad chilena. *Formación Universitaria*, 13(5), 223–232.