

**APLICACIÓN DE UN MODELO MATEMÁTICO PARA ANÁLISIS DE DATOS EN LA  
EVALUACIÓN DEL APRENDIZAJE Y LA GESTIÓN DEL CONOCIMIENTO EN UN  
INSTITUTO SUPERIOR UNIVERSITARIO**  
**APPLICATION OF A MATHEMATICAL MODEL FOR DATA ANALYSIS IN LEARNING  
ASSESSMENT AND KNOWLEDGE MANAGEMENT IN A HIGHER EDUCATION  
INSTITUTE**

**Autores:** <sup>1</sup>Dayana Cristina Villarreal Meza y <sup>2</sup> Ciro Diego Radicelli García.

<sup>1</sup>ORCID ID: <https://orcid.org/0000-0002-6971-6950>

<sup>2</sup>ORCID ID: <https://orcid.org/0000-0001-9188-0514>

<sup>1</sup>E-mail de contacto: [cristina.villarreal@unach.edu.ec](mailto:cristina.villarreal@unach.edu.ec)

<sup>2</sup>E-mail de contacto: [cradicelli@unach.edu.ec](mailto:cradicelli@unach.edu.ec)

Afiliación: <sup>1</sup>\*<sup>2</sup>Universidad Nacional del Chimborazo, (Ecuador).

Artículo recibido: 13 de mayo del 2025

Artículo revisado: 15 de mayo del 2025

Artículo aprobado: 6 de junio del 2025

<sup>1</sup>Ingeniería Industrial graduada de la Escuela Superior Politécnica del Chimborazo, (Ecuador) con 10 años de experiencia laboral. Magíster en Gestión de Operaciones, graduada de la Universidad Técnica de Ambato, (Ecuador). Candidata a Magíster en Matemática Aplicada con mención en Matemática Computacional, Universidad Nacional del Chimborazo, (Ecuador).

<sup>2</sup>Ingeniero en Sistemas Informáticos de la Escuela Superior Politécnica del Chimborazo, (Ecuador) con 20 años de experiencia laboral. Magíster en Interconectividad de Redes de la Escuela Superior Politécnica del Chimborazo, (Ecuador); Máster Universitario en Tecnologías, Sistemas y Redes de Comunicaciones de la Universidad Politécnica de Valencia, (España). Doctor dentro del Programa en Telecomunicación de la Universidad Politécnica de Valencia, (España).

### **Resumen**

El presente estudio abordó la necesidad de comprender los factores que influyen en el rendimiento académico de estudiantes del primer nivel de la carrera de Contabilidad, jornada nocturna, en un instituto superior universitario. Se planteó como objetivo principal construir y comparar modelos matemáticos que permitan identificar patrones predictivos asociados a la aprobación o reprobación estudiantil. Se aplicó una metodología cuantitativa, de alcance explicativo y enfoque correlacional, utilizando técnicas de regresión logística binaria sobre una muestra intencionada. Inicialmente, se desarrolló un modelo orientado a la evaluación del aprendizaje, que ofreció un ajuste aceptable pero limitado en términos de poder predictivo. Posteriormente, se estimó un modelo centrado en la gestión del conocimiento, del cual se seleccionó una versión reducida como resultado óptimo. Este modelo mostró métricas destacables: un AIC bajo (30.061), un área bajo la curva ROC (AUC) de 0.9883 y un McFadden  $R^2$  de 0.8496. Además, se identificó que la variable “ha repetido al menos una materia” tuvo un efecto negativo altamente significativo sobre la probabilidad de aprobación. Estos

hallazgos confirmaron la importancia de la trayectoria académica previa como indicador de riesgo, coincidiendo con investigaciones previas sobre factores de bajo desempeño. Se concluyó que el modelo reducido constituye una herramienta predictiva sólida, con potencial para apoyar la toma de decisiones institucionales orientadas a fortalecer el acompañamiento académico y prevenir la deserción estudiantil.

**Palabras clave:** Rendimiento académico, Regresión logística, Gestión del conocimiento, Estudiantes de nivel superior, Modelos estadísticos.

### **Abstract**

This study addressed the need to understand the factors influencing academic performance among first-level students of the Accounting program in the evening shift at a higher education institute. The main objective was to construct and compare mathematical models capable of identifying predictive patterns associated with student approval or failure. A quantitative methodology with an explanatory scope and correlational approach was applied, using binary logistic regression techniques on a purposive sample. Initially, a model focused on learning evaluation was developed, which

yielded acceptable adjustment but limited predictive power. Subsequently, a model centered on knowledge management was estimated, from which a reduced version was selected as the optimal result. This model presented outstanding metrics: a low AIC (30.061), a high area under the ROC curve (AUC = 0.9883), and a McFadden  $R^2$  of 0.8496. Furthermore, the variable "has repeated at least one subject" showed a highly significant negative effect on the probability of approval. These findings confirmed the importance of academic trajectory as a risk indicator, aligning with previous studies on low-performance factors. It was concluded that the reduced model constitutes a robust predictive tool, with potential to support institutional decision-making aimed at strengthening academic support and preventing student dropout.

**Keywords: Academic performance, Logistic regression, Knowledge management, Higher Education students, Statistical models.**

### **Sumário**

O presente estudo abordou a necessidade de compreender os fatores que influenciam o desempenho acadêmico de estudantes do primeiro nível do curso de Contabilidade, no turno noturno, em um instituto superior universitário. O principal objetivo foi construir e comparar modelos matemáticos que permitissem identificar padrões preditivos associados à aprovação ou reprovação estudantil. Aplicou-se uma metodologia quantitativa, de alcance explicativo e enfoque correlacional, utilizando técnicas de regressão logística binária sobre uma amostra intencional. Inicialmente, foi desenvolvido um modelo voltado para a avaliação da aprendizagem, o qual apresentou um ajuste aceitável, porém limitado em termos de poder preditivo. Posteriormente, estimou-se um modelo centrado na gestão do conhecimento, do qual se selecionou uma versão reduzida como resultado ideal. Este modelo apresentou métricas destacadas: um AIC baixo (30.061), uma área sob a curva ROC (AUC) de 0,9883 e um  $R^2$  de McFadden de 0,8496. Ademais,

identificou-se que a variável "repetiu pelo menos uma disciplina" teve um efeito negativo altamente significativo sobre a probabilidade de aprovação. Esses achados confirmaram a importância da trajetória acadêmica anterior como indicador de risco, corroborando investigações anteriores sobre fatores de baixo desempenho. Concluiu-se que o modelo reduzido constitui uma ferramenta preditiva robusta, com potencial para apoiar a tomada de decisões institucionais voltadas ao fortalecimento do acompanhamento acadêmico e à prevenção da evasão estudantil.

**Palavras-chave: Desempenho acadêmico, Regressão logística, Gestão do conhecimento, Estudantes do ensino superior, Modelos estatísticos.**

### **Introducción**

Según Bautista et al. (2024) las instituciones de educación superior enfrentan actualmente el desafío de optimizar sus procesos de enseñanza aprendizaje mediante un uso eficiente de los datos académicos generados. Esta necesidad impulsa el desarrollo e implementación de herramientas matemáticas y analíticas que permiten comprender el comportamiento de los estudiantes, fortalecer la toma de decisiones pedagógicas y consolidar una gestión más efectiva del conocimiento. Solo a través de esta formación es posible garantizar que la información recolectada sea comprendida, interpretada y aplicada adecuadamente dentro de un sistema educativo que aspire a la mejora continua. En este contexto, el uso de modelos matemáticos para analizar el aprendizaje y evaluar el entorno académico cobra especial relevancia, ya que proporciona a las instituciones bases sólidas para rediseñar estrategias pedagógicas, personalizar la atención al estudiante y fortalecer la gestión del conocimiento en función de evidencias reales. De acuerdo con Domínguez et al. (2020) la analítica del aprendizaje cobra fuerza como un campo emergente que permite recolectar,

procesar e interpretar grandes volúmenes de información educativa con fines de mejora institucional. Este enfoque se fundamenta en la idea de que cada acción del estudiante en entornos digitales deja una huella que puede transformarse en conocimiento útil para rediseñar prácticas pedagógicas y optimizar los procesos de enseñanza. No obstante, los autores advierten que este tipo de análisis debe incorporar dimensiones éticas, pedagógicas y culturales, pues el uso de algoritmos y datos masivos en educación conlleva riesgos vinculados al sesgo, la privacidad y la equidad. Por ello, proponen que las instituciones de educación superior asuman un enfoque crítico y multidisciplinar, donde la gestión de la información se oriente no solo a la eficiencia, sino también al bienestar de los estudiantes, a la transparencia de los procesos y a la mejora estructural del sistema educativo.

Lo mencionado por Suazo, (2023); Aymerich & Albarracín (2022) señala que los modelos matemáticos en los análisis estadísticos aplicados a la educación superior, permiten identificar patrones de comportamiento, predecir resultados académicos y clasificar perfiles estudiantiles a partir de la gestión eficiente del conocimiento institucional. Esta perspectiva resulta especialmente pertinente en contextos educativos donde convergen diversas modalidades de estudio, como sucede en el Instituto Superior Universitario Carlos Cisneros, que demanda una sistematización efectiva de sus datos académicos para la toma de decisiones pedagógicas fundamentadas en evidencia. Además, estos modelos permiten establecer vínculos entre las estrategias de enseñanza y los resultados obtenidos, facilitando ajustes curriculares, mejoras en los recursos didácticos y un seguimiento más preciso del desempeño estudiantil a lo largo del proceso formativo.

Desde un enfoque científico y educativo de Legusov et al., (2022; Tavares (2021); Van der Velden et al. (2023), resulta fundamental comprender cómo los estudiantes perciben las políticas institucionales relacionadas con su trayectoria académica y el acompañamiento que reciben a lo largo de su formación, ya que estas inciden directamente en la equidad del aprendizaje, la retención estudiantil y el bienestar dentro del entorno educativo. Diversos estudios han demostrado que atender las particularidades del recorrido académico de cada estudiante, reconociendo sus fortalezas y dificultades, no solo favorece un entorno más inclusivo y humano, sino que también contribuye a cerrar las brechas de rendimiento y fortalecer los procesos de gestión del conocimiento. Según Devine (2022) muchas instituciones de educación superior cuentan con normativas orientadas al fortalecimiento del aprendizaje y la permanencia estudiantil, su implementación suele verse limitada por la falta de recursos o de personal capacitado. Esta realidad afecta especialmente la eficacia de los mecanismos de seguimiento académico y apoyo a estudiantes en riesgo. Como se ha evidenciado los estudios previos citados, incluso los programas mejor diseñados pueden perder efectividad si no se respaldan con capacidades técnicas adecuadas. En este contexto, el modelo predictivo propuesto en esta investigación representa una alternativa técnica valiosa para compensar estas deficiencias, facilitando decisiones académicas más oportunas, objetivas y personalizadas.

Desde una visión humanista y transdisciplinaria, la gestión del conocimiento en contextos universitarios no debe reducirse únicamente a procedimientos técnicos, sino que implica una dimensión profundamente vinculada con el talento humano, la cultura organizacional y la transformación social. En

esta línea, Díaz (2024); Mendoza y Bullón (2022) señala que el conocimiento no se recibe de forma pasiva, sino que es una construcción activa, situada y contextual, donde el docente y el entorno juegan un papel clave en su desarrollo y circulación. Esta concepción se relaciona con los hallazgos del presente estudio, en tanto se evidencia la necesidad de implementar modelos predictivos que reconozcan no solo los indicadores académicos tradicionales, sino también las realidades institucionales, humanas y organizativas que afectan el aprendizaje. Asimismo, Cuadrado (2020); Ocaña et al. (2023) indican que los procesos de evaluación de la gestión del conocimiento en las universidades permiten identificar y valorar el desempeño organizacional mediante herramientas específicas que integran tanto los procesos internos como los resultados tangibles de conocimiento. Los autores de este artículo plantean que la gestión del conocimiento debe entenderse como un proceso organizacional complejo, dinámico y continuo, capaz de articular dimensiones como la creación, transferencia, almacenamiento y aplicación del conocimiento dentro del entorno académico.

Los planteamientos metodológicos propuestos por Ruipérez (2020; Devia (2021), proponen estudios sobre la importancia de modelos matemáticos predictivos aplicados a la evaluación del aprendizaje, como parte de una estrategia práctica en el marco de las analíticas de aprendizaje. En esa misma línea, lo argumentado por Campos et al. (2022) enfatiza que los enfoques analíticos aplicados al ámbito educativo permiten no solo recopilar e interpretar datos de forma sistemática, sino también orientar intervenciones pedagógicas más eficaces, personalizadas y contextualizadas. Los autores enfatizan que esta tecnología, bien aplicada, posibilita ajustar de

manera inmediata los procesos de enseñanza y evaluación, brindando a las instituciones una herramienta poderosa para la toma de decisiones educativas fundamentadas en evidencia empírica. Como señala Mora (2015) el rendimiento académico es el resultado de una combinación compleja de factores personales, pedagógicos e institucionales, que trascienden las simples calificaciones obtenidas por los estudiantes. Entre los elementos más influyentes se encuentran la motivación, el autoconcepto, el contexto familiar y las metodologías docentes empleadas, lo cual demuestra que las evaluaciones académicas deben interpretarse dentro de un marco más amplio de análisis integral. Esta visión resulta clave para instituciones como el Instituto Superior Universitario Carlos Cisneros, donde se requiere comprender en profundidad las múltiples variables que inciden en el proceso de aprendizaje y desempeño estudiantil.

En sintonía con esta postura, lo indicado por Mora et al. (2020) sostiene que los estudiantes que desarrollan habilidades de autorregulación y cuentan con apoyo institucional tienden a lograr aprendizajes más profundos y sostenibles. Estos procesos, que incluyen la planificación, el control metacognitivo y la autorreflexión sobre el propio desempeño, resultan especialmente importantes en contextos de educación virtual, donde el acompañamiento institucional puede marcar una diferencia significativa en los resultados académicos. Andrade et al. (2022) recomiendan ampliar la perspectiva del análisis académico, incluyendo variables que reflejan de forma más completa las condiciones reales de aprendizaje y no solo los resultados cuantitativos. Lo planteado por Acuña et al. (2021; Escorcía y Guzmán (2020); Huaranga y Méndez (2022) subrayan que la gestión del conocimiento se convierte en un componente clave cuando las

instituciones integran sus datos académicos en procesos de mejora pedagógica y toma de decisiones estratégicas. Los autores acentúan que la implementación de estrategias innovadoras en la gestión del conocimiento docente, como la creación de comunidades de práctica y repositorios institucionales, fortalece la colaboración académica y optimiza la producción científica, contribuyendo así a la mejora de la calidad educativa. En ese mismo sentido, Sánchez y Martínez (2020); Martínez (2021) proponen integrar instrumentos de evaluación que no solo midan el rendimiento, sino que valoren aspectos como la retroalimentación, la motivación y la participación activa. En este contexto, la presente investigación propone la aplicación de un modelo matemático que no solo recoja y transforme datos, sino que los analice, con el fin de evaluar cómo variables del entorno educativo como la asistencia, parámetros de evaluación, aprobación, factores sociales y tiempo de estudio influyen en el rendimiento de los estudiantes. Además, se busca identificar cómo esta información puede ser utilizada por la institución para retroalimentar sus procesos de enseñanza-aprendizaje, contribuyendo a una gestión del conocimiento más ética, precisa y sostenida en el Instituto Superior Universitario Carlos Cisneros, de la ciudad de Riobamba en Ecuador.

### **Materiales y Métodos**

La presente investigación se desarrolla bajo un enfoque metodológico cuantitativo, sustentado en el análisis de datos académicos registrados en el Instituto Superior Universitario Carlos Cisneros. La población objeto de estudio está conformada por 120 estudiantes de primer nivel de la carrera de Contabilidad, quienes cursan su formación bajo la modalidad presencial. En su defecto, la muestra corresponde al total de la población, considerando un muestreo no

probabilístico por conveniencia, justificado por la disponibilidad inmediata de los registros académicos completos, organizados y verificados por la unidad académica correspondiente, lo que facilita un análisis inicial riguroso y viable. Este tipo de muestra según Hernández y Mendoza (2018) corresponde a un subgrupo de la población en el que la elección de los elementos no depende de la probabilidad, sino de las características específicas de la investigación, como la accesibilidad a los datos, el contexto institucional y la pertinencia de los estudiantes seleccionados para los objetivos del estudio.

El estudio se centra en variables académicas institucionalmente registradas, tales como promedio, porcentaje de asistencia, calificaciones finales de 5 asignaturas, repetido al menos una materia, nivel de formación de la madre, nivel de formación del padre, cantidad de miembros en el hogar, ingreso total monetario en el hogar, entre otros factores sociales. La recolección de datos se realiza a partir de las bases documentales internas del sistema académico, garantizando su validez mediante procedimientos de depuración y verificación estructural. El diseño de la investigación, según Bastidas (2009) es de tipo descriptivo y transversal, ya que se examina la información recogida en un momento determinado, sin manipulación de las variables y sin intervenir en los procesos educativos o modificar las condiciones preexistentes. Para el análisis de los datos, se aplicaron técnicas estadísticas como la normalización de variables con el fin de unificar escalas y mejorar la precisión del modelo, además el modelo de regresión lineal múltiple, lo cual permitió identificar patrones asociados a la evaluación del aprendizaje y explorar relaciones significativas entre las variables estudiadas.

Se aplicaron herramientas del análisis estadístico multivariante, entendido como “la parte de la estadística y del análisis de datos que estudia, analiza, representa e interpreta los datos que resultan de observar más de una variable estadística sobre una muestra de individuos” (Cuadras, 2019, p. 31). En este estudio, se empleó específicamente la regresión lineal múltiple como técnica principal, con el propósito de identificar relaciones significativas entre variables académicas, personales y contextuales que influyen en la evaluación del aprendizaje y la gestión del conocimiento. Las variables consideradas fueron observadas directamente (como la asistencia, el nivel formativo de los padres, el ingreso familiar y la trayectoria académica), y fueron normalizadas para garantizar la comparabilidad de escalas. La validación del modelo se realizó mediante pruebas de significancia, el coeficiente de determinación ( $R^2$ ) y el análisis de residuos, además se utilizó la matriz de confusión, p-valor, lo cual permitió interpretar patrones de comportamiento estudiantil a partir de datos institucionales.

En su conjunto, el modelo matemático desarrollado tuvo como propósito establecer relaciones entre las variables que inciden en el rendimiento académico y la generación de evidencia empírica útil para retroalimentar la práctica docente, optimizar la planificación educativa y fortalecer la gestión del conocimiento institucional en el Instituto Superior Universitario Carlos Cisneros. La evaluación del aprendizaje, se trabajó a través de un conjunto de variables que permitieron analizar el rendimiento académico, las mismas se clasifican en cuatro categorías según su naturaleza estadística; variables cuantitativas continuas; asistencia (expresada como porcentaje de asistencia de 0 a 100), ingreso Total Hogar (medido en dólares), calificaciones

académicas en asignaturas como matemática básica, expresión oral escrita, inglés\_1, administración, contabilidad básica y ofimática básica, con escalas del 0 al 10. Variables cuantitativas discretas; cantidad Miembros Hogar (número total de integrantes del hogar). Variables categóricas ordinal, nivel formación madre y nivel formación padre (codificados según el máximo nivel educativo alcanzado). Variables categóricas dicotómicas; ha repetido al menos una materia (1 = sí ha repetido, 0 = no ha repetido).

Para asegurar la comparabilidad y robustez del análisis multivariado, se procedió a normalizar únicamente las variables cuantitativas continuas. Esta transformación se realizó aplicando la estandarización Z (media cero y desviación estándar uno), de acuerdo con los criterios establecidos en la literatura especializada. Según Hair et al. (2019), en los estudios de análisis multivariante es indispensable normalizar aquellas variables que se expresan en escalas diferentes para evitar que dominen los resultados numéricamente y para facilitar la interpretación comparativa de los coeficientes dentro del modelo. En este sentido, se estandarizaron las variables asistencia, ingreso total hogar y cantidad de miembros de hogar, dado que presentan unidades distintas a las calificaciones y demás factores del modelo. Por otro lado, variables categóricas como “ha repetido al menos una materia”, ya codificadas como binarias (0 y 1), no fueron sometidas a transformación, pues la normalización en este caso distorsionaría su significado lógico. La Tabla 1 presenta el resumen estadístico descriptivos, de las variables involucradas en el modelo de estudio, ya transformadas mediante estandarización z (z-score), lo cual permite su comparación en una misma escala. Entre las variables analizadas se incluyen las calificaciones por asignatura, la asistencia, y

factores contextuales como el nivel educativo de los padres, el ingreso del hogar y el tamaño del núcleo familiar.

Como se observa, la variable (Asistencia) tiene una media cercana a 0 y una desviación estándar de aproximadamente 1, confirmando la correcta normalización. Asimismo, las demás variables numéricas muestran una dispersión acorde a lo esperado bajo estandarización. Por ejemplo, la variable (ingreso total hogar) oscila entre -1.63 y 5.33 en su forma estandarizada, reflejando la desigualdad presente en los recursos económicos de los estudiantes. La variable categórica dicotómica (ha repetido al menos una materia) fue codificada como binaria (0 = No ha repetido, 1 = Ha repetido), manteniéndose en su escala natural sin ser normalizadas para no perder su significado interpretativo. Según Rencher y Christensen (2002), la estandarización de variables cuantitativas es una etapa clave en análisis multivariado, especialmente cuando se pretende aplicar técnicas como regresión múltiple, donde se requiere homogeneidad en la escala de medición.

Para analizar los factores que inciden en la evaluación del aprendizaje de los estudiantes, se estimó un modelo de regresión lineal múltiple

**Tabla 1.** Resumen descriptivo de variables en estudio para el modelo evaluación del aprendizaje

Variable	Mínimo	Q1	Mediana	Media	Q3	Máximo
Asistencia	-3.8387	0.2134	0.4693	0.4601	0.5316	0.5593
Matemática Básica	0.0000	7.0000	9.0000	7.2120	9.3000	10.0000
Expresión Oral Escrita	0.0000	7.0000	8.5900	7.4590	9.5600	10.0000
Inglés I	0.0000	7.0000	9.0000	7.9020	9.6900	10.0000
Administración Básica	0.0000	7.7470	9.1600	7.7530	9.6000	10.0000
Contabilidad Básica	0.0000	7.2370	8.4450	7.3680	9.4630	10.0000
Ofimática Básica	0.0000	7.3550	9.0800	7.3260	9.8000	10.0000
Ingreso Total del Hogar	-1.6337	-0.5884	0.0309	0.0000	0.2129	5.3348
Cantidad Miembros del Hogar	-2.2397	-0.6009	-0.0546	0.0000	0.4916	3.7692
Ha repetido al menos una materia	1.0000	2.0000	2.0000	1.7750	2.0000	2.0000

**Fuente:** elaboración propia

en el que la variable dependiente fue el promedio general de calificaciones (promedio), calculado a partir de las asignaturas cursadas por cada estudiante. Las variables independientes incluidas en el modelo fueron: asistencia (expresada en porcentaje y previamente normalizada), ingreso total del hogar, cantidad de miembros del hogar, nivel de formación de la madre, nivel de formación del padre y una variable dicotómica denominada (*ha repetido al menos una materia*), que indica si el estudiante ha repetido alguna asignatura (1 = sí, 0 = no). Estas variables fueron seleccionadas por su respaldo teórico en investigaciones sobre aprendizaje y por la disponibilidad de información confiable en la base de datos utilizada. El modelo fue ajustado con variables normalizadas, a fin de garantizar la comparabilidad entre diferentes escalas de medición, conforme a las recomendaciones metodológicas propuestas por Hair et al. (2019).

### **Resultados y Discusión**

En base a lo anteriormente mencionado, la estructura del modelo quedó representada de la siguiente forma: promedio ~ asistencia + ingreso total hogar + cantidad miembros Hogar + nivel formación Madre + nivel formación Padre + ha repetido al menos una materia.

### **Análisis del Modelo de Evaluación del Aprendizaje**

El modelo tuvo como objetivo identificar los factores que inciden en el rendimiento académico de los estudiantes, operacionalizado a través de la variable dependiente promedio. La especificación incluyó variables académicas y sociodemográficas, entre ellas: asistencia, nivel de formación de los padres, ingreso total del hogar, cantidad de miembros del hogar y la variable dicotómica (ha repetido al menos una materia). Los resultados estadísticos de la tabla 2, evidenciaron un coeficiente de determinación ( $R^2$ ) de 0.9109, lo cual indica que el 91.09% de la variabilidad observada en el promedio de los estudiantes puede ser explicada por el conjunto de variables independientes consideradas. A su vez, el  $R^2$  ajustado de 0.8907 confirma que, incluso considerando el número de predictores, el modelo mantiene un alto poder explicativo. La prueba F resultó significativa ( $F(22,97) = 45.09, p < 0.001$ ), lo que valida la pertinencia global del modelo.

De manera particular, se encontró que la (asistencia a clases) *es el predictor más relevante* en términos estadísticos y prácticos ( $\beta = 1.1195, p < 0.001$ ), lo que sugiere que, a mayor porcentaje de asistencia, mayor es el promedio académico alcanzado por los estudiantes. Este resultado coincide con investigaciones previas que destacan la influencia directa de la presencia regular en actividades de aula sobre el proceso de aprendizaje y rendimiento académico. Otra variable estadísticamente significativa fue (ha repetido al menos una materia) ( $\beta = 4.2828, p < 0.001$ ), el signo positivo del coeficiente indica que, en promedio, los estudiantes que han repetido al menos una materia presentan una calificación promedio más alta que aquellos que no lo han hecho, una vez controladas las demás variables del modelo. Este hallazgo podría

parecer contraintuitivo, ya que la repetición suele asociarse con bajo desempeño académico. No obstante, este resultado podría estar reflejando un efecto de recuperación o compensación académica, donde los estudiantes que han repetido una materia desarrollan posteriormente un mayor esfuerzo, motivación o compromiso académico, que se traduce en mejores resultados en periodos siguientes. También es posible que algunos de estos estudiantes cuenten con apoyo institucional o familiar tras haber repetido, lo cual influye en su desempeño posterior.

En cuanto al nivel educativo de los padres, se identificó que el nivel de formación del padre en educación básica (código 4) tiene un efecto negativo y estadísticamente significativo sobre el rendimiento académico ( $\beta = -1.3057, p = 0.0495$ ). Este hallazgo resalta la importancia del entorno educativo familiar en el desarrollo académico del estudiante, y sugiere que una menor formación del padre podría limitar el acompañamiento o la motivación hacia el logro escolar. Este hallazgo resalta la importancia del entorno educativo familiar en el desarrollo académico del estudiante, y sugiere que una menor formación del padre podría limitar el acompañamiento o la motivación hacia el logro escolar. En este contexto, se vuelve fundamental que el Instituto Superior Universitario identifique estos perfiles mediante el análisis de datos institucionales, y que gestione estrategias de intervención orientadas a fortalecer el conocimiento en los estudiantes. Estas acciones pueden incluir programas de tutoría académica, orientación vocacional o acompañamiento psicológico, lo cual se alinea con un enfoque de gestión del conocimiento institucional, en el que la información se transforma en decisiones pedagógicas y apoyos efectivos.

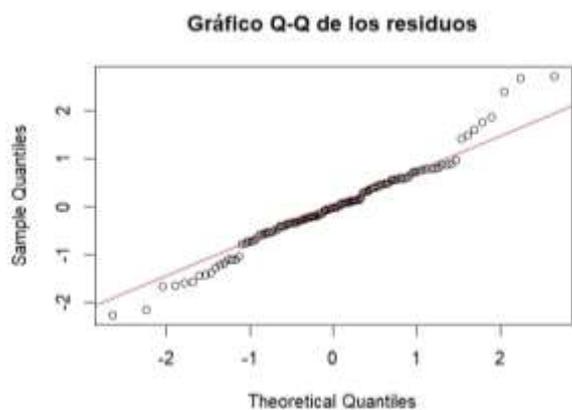
Otras variables, como el ingreso total del hogar y la cantidad de miembros, si bien conceptualmente relevantes, no resultaron significativas en este modelo. Asimismo, la mayoría de los niveles del nivel de formación de la madre y del padre (excepto el mencionado anteriormente) no mostraron efectos estadísticamente significativos, lo cual puede atribuirse a la dispersión de frecuencias dentro de cada categoría o a la existencia de multicolinealidad entre variables dummy, es decir, variables categóricas codificadas numéricamente que solo toman los valores 0 y 1 para representar dos estados opuestos o mutuamente excluyentes. En relación con los residuos del modelo, se observa que presentan una distribución centrada en torno a cero, con una mediana de -0.009. Este valor cercano a cero sugiere que no existe un sesgo sistemático en los errores de predicción. Asimismo, los

valores correspondientes al primer y tercer cuartil (Q1 = -0.463; Q3 = 0.516) indican que el 50% central de los residuos se encuentra contenido dentro de un rango moderado, lo que refleja una variabilidad controlada en los errores del modelo. El rango de los residuos oscila entre -2.263 y 2.716, lo cual resulta aceptable dado que las variables del modelo fueron previamente estandarizadas. Esta dispersión no evidencia la presencia de valores atípicos extremos que puedan afectar la estabilidad del modelo. Estos resultados son coherentes con el gráfico Q-Q de los residuos, en el cual los puntos se alinean de forma general sobre la línea de referencia, lo que sugiere que la distribución de los residuos es normal. Esta normalidad es fundamental para garantizar la validez inferencial del modelo de regresión lineal, tal como se establece en la literatura metodológica Gareth & Trevor (2021).

**Tabla 2. Resultados del modelo lineal múltiple evaluación del aprendizaje**

Coeficientes: (1 no definido debido a singularidades)					
	Estimación	Error Estándar	Valor t	p-valor	Significancia
(Intercepto)	-0.0137	0.97833	-0.014	0.9889	
ASISTENCIA	1.1950	0.20599	5.435	4.08e-07	***
Ingreso Total Hogar	-0.0489	0.10354	-0.473	0.6374	
Cantidad Miembros Hogar	-0.1062	0.10115	-1.050	0.2962	
Nivel de formación de la Madre1	1.0498	1.11593	0.940	0.3653	
Nivel de formación de la Madre10	-0.3961	1.02892	-0.385	0.7007	
Nivel de formación de la Madre2	0.39661	1.12789	0.351	0.7257	
Nivel de formación de la Madre 3	0.61791	0.51833	1.192	0.2361	
Nivel de formación de la Madre 4	0.52708	0.60974	0.865	0.3872	
Nivel de formación de la Madre 5	0.54159	0.51172	1.058	0.2925	
Nivel de formación de la Madre 6	0.42650	0.73888	0.576	0.5659	
Nivel de formación de la Madre 7	1.00300	0.73361	1.404	0.1633	
Nivel de formación de la Madre 8	0.40104	0.91774	0.437	0.6631	
Nivel de formación de la Madre 9	0.64904	1.13303	0.573	0.5681	
Nivel de formación del Padre1	-1.2049	1.28270	-0.939	0.3501	
Nivel de formación del Padre 10	-0.61510	0.58260	-1.056	0.2937	
Nivel de formación del Padre 2	NA	NA	NA	NA	
Nivel de formación del Padre 3	-0.85014	0.44871	-1.895	0.0611	.
Nivel de formación del Padre 4	-1.5071	0.65627	-1.990	0.0495	*
Nivel de formación del Padre 5	-1.32484	0.47035	-1.116	0.2642	
Nivel de formación del Padre 6	0.18249	0.72774	0.251	0.8025	
Nivel de formación del Padre 7	-0.37344	0.59682	-0.626	0.5330	
Nivel de formación del Padre 8	-0.68848	0.62929	-1.094	0.2766	
Ha Repetido Al Menos Una Materia	4.28283	0.47979	8.926	2.78e-14	***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 0.9655 on 97 degrees of freedom			Multiple R-squared: 0.9109,		Adjusted R-squared: 0.8907
F-statistic: 45.09 on 22 and 97 DF, p-value: < 2.2e-16					
Residuals:	Min	1Q	Median	3Q	Max
	-2.26295	-0.46309	-0.00937	0.51574	2.71552

**Fuente:** elaboración propia



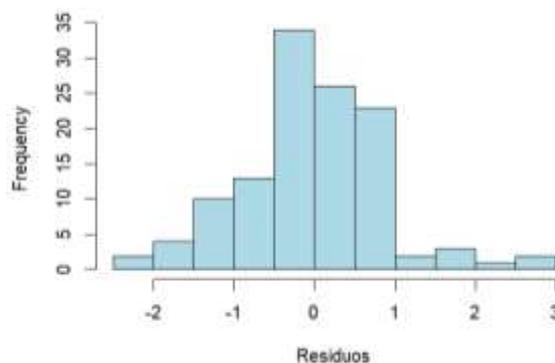
**Figura 1.** *Dispersión de residuos del modelo de evaluación del aprendizaje*

La figura 1 de residuos frente a los valores ajustados muestra una dispersión aleatoria alrededor de la línea cero, sin evidencias visibles de patrones sistemáticos. Esta distribución sugiere que se cumple el supuesto de homocedasticidad en el modelo de regresión, es decir, los errores presentan varianza constante a lo largo del rango de predicción. Por tanto, se considera que el modelo mantiene una adecuada estabilidad en sus residuos.

### **Análisis de los resultados del test de Shapiro-Wilk**

Para verificar el cumplimiento del supuesto de normalidad de los residuos, se aplicó la prueba de Shapiro-Wilk, debido a que la muestra es mayor a 50 individuos. El resultado obtenido fue  $W = 0.9716$  con un valor  $p = 0.0122$ , lo que indica una diferencia estadísticamente significativa respecto a una distribución normal. En términos prácticos, este resultado sugiere que los residuos no siguen una distribución normal perfecta. Sin embargo, considerando el tamaño de la muestra y que la desviación observada en el gráfico Q-Q no presenta distorsiones severas, esta leve violación de la normalidad puede considerarse tolerable, especialmente si el modelo es utilizado con fines descriptivos o explicativos. De acuerdo con Tabachnick y Fidell (2013), en muestras moderadas, pequeñas desviaciones de la normalidad no afectan

sustancialmente la validez de los resultados de la regresión lineal, siempre que no existan otros problemas como heterocedasticidad o colinealidad severa.



**Figura 2.** *Histograma de los residuos del modelo de regresión lineal múltiple, evaluación del aprendizaje*

### **Análisis del histograma de los residuos**

El histograma de los residuos que se observa en la figura 2, muestra una distribución aproximadamente simétrica, con una concentración central alrededor del valor cero. Aunque se observan ligeras asimetrías y algunas frecuencias dispersas en los extremos (particularmente en el intervalo derecho), la forma general de la distribución es compatible con una curva normal moderadamente aceptable. Este resultado, en conjunto con el gráfico Q-Q y el test de Shapiro-Wilk, sugiere que la normalidad de los errores no se cumple estrictamente, pero las desviaciones no son lo suficientemente severas como para invalidar el modelo, especialmente considerando el enfoque exploratorio y descriptivo de esta investigación. De acuerdo con Cuadras (2019), en análisis multivariados aplicados, es posible tolerar ligeras violaciones al supuesto de normalidad, siempre que no existen otros problemas estructurales en el modelo, en este caso más aún al campo educativo. Esta flexibilidad resulta aún más pertinente en investigaciones del ámbito educativo, donde las condiciones ideales de

distribución suelen ser difíciles de cumplir plenamente.

### **Modelo para gestión del conocimiento**

En el modelo de gestión del conocimiento se utilizaron distintas tipologías de variables clasificadas según su naturaleza estadística, de acuerdo con los lineamientos metodológicos comúnmente aceptados en el análisis de modelos logísticos. A continuación, se detallan las variables empleadas y su respectiva clasificación. En primer lugar, se consideró una variable cuantitativa continua: edad, expresada en años completos cumplidos al momento de la recolección de datos. Esta variable es numérica y puede asumir cualquier valor dentro de un rango definido, permitiendo su tratamiento mediante operaciones aritméticas y su inclusión directa en modelos de regresión. En segundo lugar, se incorporaron variables categóricas dicotómicas, codificadas de forma binaria para su análisis en el modelo logístico.

Estas incluyen: ha repetido al menos una materia (1 = ha repetido, 2 = no ha repetido), que evalúa si el estudiante ha tenido historial de repetición académica; ha perdido la gratuidad (1 = sí ha perdido, 2 = no ha perdido), relacionada con la pérdida del derecho a gratuidad educativa según lo dispuesto en el artículo 80 de la LOES; género (1 = masculino, 2 = femenino), que identifica el sexo del estudiante; y estudiante ocupación (1 = solo estudia, 2 = estudia y trabaja), variable que distingue entre dedicación exclusiva al estudio y combinación con actividades laborales. Cabe señalar que, si bien algunas variables fueron tratadas originalmente como cualitativas, su transformación en factores binarios permitió su incorporación eficiente en la estimación del modelo logístico binario, garantizando así la interpretabilidad de los coeficientes y la consistencia de los resultados estadísticos.

### **Análisis descriptivo de las variables del modelo gestión del conocimiento**

A continuación, se establece el análisis descriptivo de la variable cuantitativa edad en el modelo matemático de gestión del conocimiento, revela patrones relevantes dentro de la población estudiantil analizada presenta una media de 30.55 años y una desviación estándar de 8.23, con una asimetría positiva de 1.07 y una curtosis de 0.72, lo que indica una ligera concentración de estudiantes jóvenes y una distribución sin colas extremas. En la tabla 3 se observa los resultados del modelo logístico de Gestión del conocimiento

**Tabla 3.** Resultados del modelo logístico de Gestión del conocimiento

	Estimate	Std. Error	z value	Pr(> z )
(Intercepto)	0.742	3.270	0.227	0.820
Edad	0.066	0.130	0.508	0.611
Género (2 = femenino o masculino)	-0.659	1.307	-0.504	0.614
Ha repetido al menos una materia (Sí)	-37.89	6034.7	-0.006	0.995
Ha perdido la gratuidad (Sí)	15.404	4313.6	0.004	0.997
Ocupación del estudiante (categoría 2)	-0.033	1.373	-0.024	0.981
(Dispersion parameter for binomial family taken to be 1)				
Null deviance: 120.097 on 119 degrees of freedom				
Residual deviance: 18.061 on 114 degrees of freedom				
AIC: 30.061				
Number of Fisher Scoring iterations: 21				

**Fuente:** elaboración propia

### **Análisis del modelo matemático logístico binario para gestión del conocimiento**

El modelo logístico binario formulado para predecir la probabilidad de que un estudiante apruebe (estado = 1) o repruebe (estado = 0) fue estimado a partir de un subconjunto de variables sociodemográficas y académicas, seleccionadas con base en criterios de parsimonia y estabilidad estadística. Este modelo incluyó las variables edad, género, ha repetido al menos una materia,

ha perdido la gratuidad y estudiante ocupación. En la tabla 3, se observa que, si bien ninguna de estas variables resultó estadísticamente significativa de manera individual ( $p > 0.05$ ), el modelo mostró un buen ajuste general y una estructura interpretable desde el punto de vista contextual. La variable *edad*, tratada como continua, presentó un coeficiente positivo ( $\beta = 0.0662$ ;  $p = 0.611$ ), lo cual sugiere una tendencia leve a que los estudiantes de mayor edad tengan mayor probabilidad de aprobar, aunque este efecto no fue significativo. Por su parte, la variable género, codificada como 1 = Masculino y 2 = Femenino, mostró un coeficiente negativo ( $\beta = -0.6595$ ;  $p = 0.614$ ), indicando que, en comparación con los hombres, las mujeres tendrían una ligera disminución en la probabilidad de aprobar, sin que esto sea estadísticamente concluyente.

La variable ha repetido al menos una materia, que distingue entre estudiantes que han reprobado asignaturas en el pasado (1 = Sí, 2 = No), registró un coeficiente notablemente negativo ( $\beta = -37.8920$ ;  $p = 0.995$ ). Aunque este valor sugiere un posible impacto adverso sobre el rendimiento académico, la ausencia de significancia estadística y el valor extremo del coeficiente indican posibles problemas de desbalance en la distribución de esta variable. La variable ha perdido la gratuidad (1 = Sí, 2 = No), vinculada a la pérdida del derecho a gratuidad educativa según el artículo 80 de la Ley Orgánica de Educación Superior (LOES), arrojó un coeficiente positivo ( $\beta = 15.4046$ ;  $p = 0.997$ ). Si bien este resultado no es estadísticamente significativo, podría interpretarse en el sentido de que estudiantes que han perdido este beneficio pueden mantener la aprobación, posiblemente debido a factores externos como el esfuerzo adicional o el apoyo familiar.

En relación con la variable estudiante ocupación, que diferencia entre quienes únicamente estudian

(1) y quienes estudian y trabajan (2), se observó un coeficiente muy cercano a cero ( $\beta = -0.0333$ ;  $p = 0.981$ ), lo cual sugiere que compatibilizar estudio y trabajo no representa una diferencia sustancial en el desempeño académico dentro de esta cohorte. Desde una perspectiva global, el modelo mostró un buen ajuste, con un Akaike Information Criterion (AIC) bajo (30.061) y una deviance residual de 18.061, lo cual indica que, aunque las variables individualmente no aporten una explicación estadísticamente significativa, el modelo en su conjunto logra capturar una estructura útil para comprender el fenómeno estudiado. Este modelo representa una alternativa metodológicamente más estable frente al modelo completo, el cual presentaba estimaciones inestables y coeficientes distorsionados debido a la baja frecuencia en algunas categorías. En la tabla 4 se observan los Odds Ratios Modelo Logístico Binario y su posterior análisis.

**Tabla 4. Odds Ratios Modelo Logístico Binario**

Variable	Odds Ratio	IC 95% Inferior	IC 95% Superior
(Intercepto)	2.101169e+00	1.077672e-03	1.027697e+03
Edad	1.068464e+00	8.569774e-01	1.500834e+00
Género (2)	5.170958e-01	2.160275e-02	6.265896e+00
Ha repetido al menos una materia (2)	3.4970e-17	NA	1.4958e+126
Ha perdido la gratuidad (2)	4.8990e+06	3.5985e-215	NA
estudiante ocupación (2)	9.672133e-01	3.673907e-02	1.339667e+01

Fuente: elaboración propia

A partir del modelo logístico binario, se calcularon las razones de probabilidades (odds ratios) junto con sus intervalos de confianza al 95%, tal como se presenta en la tabla 4. Esta información permite interpretar la magnitud y dirección del efecto de cada predictor sobre la probabilidad de aprobar (estado = 1). El

Intercepto presentó un odds ratio de 2.10, aunque con un intervalo extremadamente amplio (IC 95%: 0.001 – 1027), lo que sugiere poca precisión en su estimación y la necesidad de cautela al interpretarlo. La variable edad mostró un odds ratio de 1.06 (IC 95%: 0.86 – 1.50), indicando que, por cada año adicional de edad, la probabilidad de aprobar se incrementa levemente. Sin embargo, dado que el intervalo de confianza incluye el valor 1, este efecto no es estadísticamente significativo.

Para el género, codificado como 1 = Masculino y 2 = Femenino, el odds ratio fue de 0.51 (IC 95%: 0.02 – 6.26), lo que sugiere que las mujeres podrían tener una menor probabilidad de aprobar en comparación con los hombres. No obstante, el amplio rango del intervalo refleja alta incertidumbre en esta estimación. La variable ha repetido al menos una materia (1 = Sí, 2 = No) mostró un odds ratio extremadamente cercano a cero ( $3.49e-17$ ) y un límite superior de IC de  $1.49e+126$ , sin límite inferior definido (NA). Esto indica un posible problema de separación completa o distribución desbalanceada, lo cual afecta gravemente la estabilidad del coeficiente. Aun así, el sentido del efecto indica que quienes han repetido materias tienen una probabilidad sustancialmente menor de aprobar. En el caso de ha perdido la gratuidad, el odds ratio fue de  $4.89e+06$  con un intervalo de confianza inferior casi nulo ( $3.59e-215$ ) y sin límite superior. Esta inestabilidad estadística implica que, aunque el efecto estimado es fuerte y positivo, los resultados no deben interpretarse de forma aislada debido al desequilibrio en la variable. Finalmente, para la variable estudiante ocupación, que diferencia entre quienes solo estudian (1) y quienes estudian y trabajan (2), se obtuvo un odds ratio de 0.96 (IC 95%: 0.036 – 13.39), indicando prácticamente nula diferencia entre ambos grupos en términos de probabilidad

de aprobación, con un intervalo muy amplio que denota alta variabilidad

**Tabla 5. Diagnóstico de Colinealidad**

Variable	VIF
Edad	1.082606
Género	1.011212
Ha repetido al menos una materia	2.044689
Ha perdido la gratuidad	2.044689
Ocupación del estudiante	1.078836

Fuente: elaboración propia

Para evaluar la posible presencia de colinealidad entre los predictores del modelo logístico binario, la tabla 5 muestra los resultados del Factor de Inflación de la Varianza (VIF). Este indicador permite identificar si alguna variable independiente está altamente correlacionada con otras dentro del modelo, lo cual podría afectar la estabilidad de los coeficientes estimados. Los resultados obtenidos mostraron valores bajos para todas las variables: edad (VIF = 1.08), género (1.01), ha repetido al menos una materia (2.04), ha perdido la gratuidad (2.04) y estudiante ocupación (1.07). Dado que todos los VIF se encuentran muy por debajo del umbral de 5, se concluye que no existe colinealidad preocupante entre los predictores incluidos en el modelo. Esto respalda la consistencia estadística del modelo y garantiza que las estimaciones obtenidas para cada variable son independientes y no están afectadas por redundancias lineales. La ausencia de colinealidad fortalece la validez de las conclusiones obtenidas a partir de este modelo.

**Tabla 6. Evaluación del ajuste del modelo mediante coeficientes de Pseudo  $R^2$**

Ajuste del modelo nulo para pseudo-r <sup>2</sup>	
llh	-9.0304369
llhNull	-60.0482908
G2	102.0357078
McFadden	0.8496138
r <sup>2</sup> ML	0.5727122
r <sup>2</sup> CU	0.905593

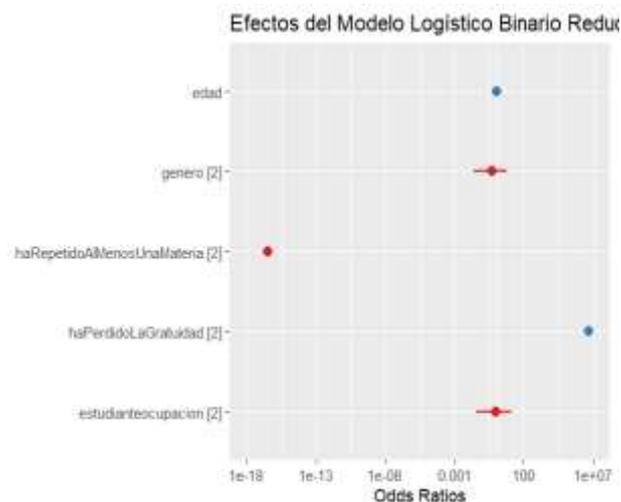
Fuente: elaboración propia

El análisis de ajuste global del modelo logístico binario, a través de los indicadores de Pseudo R<sup>2</sup> que se muestran en la tabla 6, reveló valores altamente satisfactorios: McFadden = 0.8496, Cox & Snell = 0.5727 y Nagelkerke = 0.9056. Estos resultados indican que el modelo explica una proporción significativa de la variabilidad en la probabilidad de aprobación, siendo particularmente destacable el valor de Nagelkerke, que supera el 90%. De acuerdo con criterios estadísticos, un McFadden R<sup>2</sup> superior a 0.3 ya refleja un buen ajuste, por lo que el valor alcanzado en este caso confirma la solidez predictiva y explicativa del modelo. La matriz de confusión obtenida para el modelo logístico binario evidencia una alta capacidad predictiva del modelo. De los 96 estudiantes que aprobaron, 93 fueron correctamente clasificados y solo 3 fueron clasificados incorrectamente como reprobados, mientras que los 24 estudiantes que reprobaron fueron identificados correctamente en su totalidad.

Este resultado sugiere que el modelo presenta una excelente sensibilidad para identificar casos de reprobación (100%) y una alta especificidad para detectar casos de aprobación (96.9%), lo cual refuerza su utilidad práctica como herramienta predictiva del rendimiento académico en la población estudiada. La evaluación del poder discriminativo del modelo logístico binario, mediante el cálculo del área bajo la curva ROC (AUC), arrojó un valor de 0.9883, lo que indica una capacidad predictiva excelente. Un AUC cercano a 1.0 refleja que el modelo distingue eficazmente entre los estudiantes que aprueban y los que reprueban. De acuerdo con los criterios de interpretación establecidos en la literatura estadística, un valor superior a 0.9 se considera sobresaliente, por lo que este resultado confirma la solidez del modelo como herramienta clasificatoria en el contexto del rendimiento académico.

### Efectos marginales promedio del modelo logístico binario

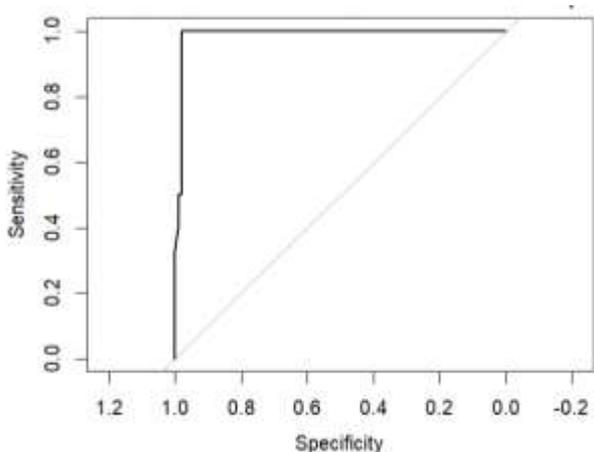
El análisis de los efectos marginales promedio del modelo logístico binario permitió cuantificar el impacto directo de cada predictor sobre la probabilidad de aprobación, manteniendo constantes las demás variables. Las cinco variables consideradas, solo una resultó estadísticamente significativa: “ha Repetido Al Menos Una Materia2”, con un efecto marginal negativo de -0.9065 ( $p < 0.001$ ), lo que indica que los estudiantes que no han repetido materias tienen aproximadamente un 91% más de probabilidad de aprobar en comparación con quienes sí han repetido al menos una asignatura. Las demás variables edad, género, ha perdido la gratuidad y estudiante ocupación no alcanzaron significancia estadística ( $p > 0.05$ ), aunque el efecto de “haPerdidoLaGratuidad2” se acercó al umbral ( $p = 0.0641$ ), lo que podría indicar una leve tendencia positiva en la probabilidad de aprobar para quienes no han perdido este beneficio.



**Figura 3.** Curva ROC - Modelo de Gestión del Conocimiento

La figura 3 muestra la curva ROC correspondiente al modelo logístico binario aplicado en el análisis del rendimiento académico. La curva evidencia una alta capacidad discriminativa, manteniéndose muy

próxima al vértice superior izquierdo del gráfico, lo que indica un excelente equilibrio entre sensibilidad (verdaderos positivos) y especificidad (verdaderos negativos). Este resultado es consistente con el valor del área bajo la curva (AUC = 0.9883), lo cual confirma que el modelo tiene una capacidad casi perfecta para diferenciar entre estudiantes que aprueban y los que reprobaban.



**Figura 4.** Efectos del Modelo Logístico Binario

La figura 4 presenta los efectos de las variables independientes sobre la probabilidad de aprobación, expresados como razones de probabilidad (odds ratios) con sus respectivos intervalos de confianza del 95%. En el gráfico se observa que la variable ha repetido al menos una materia se ubica muy por debajo del valor de referencia 1, lo que indica un efecto negativo sustancial en la probabilidad de aprobar para quienes han repetido asignaturas. En contraste, variables como edad y ha Perdido La Gratuidad muestran odds ratios elevados, aunque con rangos de incertidumbre amplios. La variable género se aproxima al valor de neutralidad ( $OR \approx 1$ ), lo que sugiere que, en esta muestra, no se identifican diferencias marcadas en el rendimiento académico entre hombres y mujeres. En conjunto, el gráfico evidencia tanto la dirección como la magnitud relativa del efecto de cada predictor, aportando una visualización clara

sobre la influencia de las variables incluidas en el modelo.

Los resultados obtenidos en esta investigación confirman que factores como la asistencia a clases y la trayectoria académica previa inciden significativamente en el rendimiento estudiantil, coincidiendo con lo expuesto por Mora (2015), quien destaca que la influencia directa de la presencia regular y el historial académico en el desempeño. Asimismo, los factores sociales, presentan una relación negativa entre la baja formación educativa del padre y el promedio académico reafirma lo planteado por Legusov et al. (2022) sobre la relevancia del entorno familiar. En contraste, variables como ingreso familiar y cantidad de miembros del hogar no mostraron significancia, lo que difiere de los hallazgos de Andrade et al. (2022), quienes subrayan el peso de los factores sociales. Estos contrastes sugieren que, en contextos institucionales específicos como el del Instituto Carlos Cisneros, ciertos determinantes del aprendizaje cobran mayor relevancia, destacando la necesidad de políticas en lo que respecta a gestión del conocimiento focalizadas basadas en evidencia. En este sentido, el acompañamiento desde la institución puede ser interpretado como parte de una estrategia de “*gestión del conocimiento*”, orientada a identificar necesidades académicas específicas y brindar respuestas pedagógicas basadas en información sistematizada. Este hallazgo merece un análisis complementario desde enfoques cualitativos que permitan comprender las dinámicas de resiliencia académica en estudiantes con trayectorias educativas irregulares.

El modelo logístico binario se consolidó como la opción más coherente y funcional. Con un AIC bajo (30.061) y una deviance residual de 18.061, presentó un buen ajuste general, pese a que pocas

variables fueron significativas de forma individual. Esta versión más parsimoniosa evitó los problemas del modelo completo, como la inestabilidad de los coeficientes por baja frecuencia en algunas categorías, lo que coincide con Cuadras (2019), quien señala que en análisis multivariados es preferible optar por modelos simples que mantengan buena capacidad explicativa, priorizando estabilidad e interpretabilidad, especialmente en contextos educativos con limitaciones estructurales. La curva ROC del modelo logístico binario mostró una excelente capacidad discriminativa, con un AUC de 0.9883. La forma de la curva, claramente alejada de la diagonal de referencia, confirma el poder del modelo para clasificar adecuadamente a los estudiantes en función de su probabilidad de aprobación o reprobación, lo cual se alinea con Hair et al. (2019), quienes destacan que la validez de un modelo predictivo no solo depende del ajuste interno, sino también de su capacidad para discriminar con precisión entre categorías relevantes, lo que refuerza su utilidad como herramienta diagnóstica para la gestión educativa basada en datos.

Uno de los hallazgos más sólidos fue el efecto de la variable “ha repetido al menos una materia”, que mostró un impacto negativo significativo. Los estudiantes que no repitieron asignaturas tuvieron un 91% más de probabilidad de aprobar, lo que refuerza el valor predictivo de la trayectoria académica y plantea la necesidad de profundizar en los factores que pueden fomentar la resiliencia estudiantil. Estos patrones podrían estar vinculados a acciones institucionales posteriores a la repetición, como tutorías o apoyos pedagógicos, orientadas a la mejora continua, lo que concuerda con Suárez (2023), quien sostiene que el aprendizaje se construye en contextos dinámicos y activos, donde el entorno y la intervención docente desempeñan un papel clave en el desarrollo académico, y resalta la

importancia de implementar estrategias de gestión del conocimiento que permitan identificar a tiempo las trayectorias educativas irregulares y generar respuestas pedagógicas efectivas basadas en información sistematizada.

### **Conclusiones**

La presente investigación permitió desarrollar y contrastar dos modelos matemáticos aplicados al análisis del rendimiento académico: el modelo de evaluación del aprendizaje y el modelo de gestión del conocimiento. El primero aportó una visión general sobre la influencia de ciertas variables académicas básicas, aunque con limitaciones en su capacidad predictiva. En cambio, el modelo logístico binario centrado en la gestión del conocimiento presentó una estructura más sólida, permitiendo identificar patrones relevantes vinculados a la trayectoria académica de los estudiantes en cuanto a factores sociales. Tras evaluar distintas configuraciones estadísticas, el modelo de gestión del conocimiento fue el que evidenció el mejor ajuste global. Mostró métricas destacables como un AIC bajo (30.061), una curva ROC con AUC de 0.9883, y un McFadden  $R^2$  de 0.8496, superando significativamente a los modelos iniciales en estabilidad, precisión y simplicidad estructural.

Estas características respaldan su validez como herramienta predictiva dentro de contextos educativos similares. Por tanto, se concluye que el modelo cumple con el supuesto de normalidad en los residuos, dado que estos se distribuyen de manera centrada, sin sesgos sistemáticos, y con una dispersión controlada. Esta condición refuerza la validez del modelo logístico binario para evaluar la relación entre las variables independientes y el desempeño académico, garantizando una base sólida para la inferencia estadística en contextos educativos. A pesar de que la mayoría de las variables incluidas en el modelo para la gestión del conocimiento, no

resultaron significativas de manera individual, el análisis de los efectos marginales reveló que haber repetido al menos una materia tiene un impacto negativo altamente significativo en la probabilidad de aprobación. Se recomienda que futuras investigaciones amplíen el conjunto de variables consideradas, incorporando dimensiones psicológicas, motivacionales y familiares, que podrían aportar mayor capacidad explicativa al modelo. Asimismo, sería pertinente aplicar este enfoque a diferentes carreras o modalidades académicas del instituto, con el objetivo de validar la generalización del modelo y fortalecer su aplicabilidad institucional.

### Referencias Bibliográficas

- Acuña, M. (2021). Manejo de emociones en estudiantes universitarios en tiempos de confinamiento: Una Propuesta de intervención desde un proyecto de aula. *Boletín de Innovación, Logística y Operaciones*, 3(1). <https://doi.org/10.17981/bilo.3.1.2021.08>
- Aymerich, À., & Albarracín, L. (2022). Mathematical Modeling in Statistical Activities: Key Episodes for Model Generation. *Uniciencia*, 36(1), 1-16. <http://dx.doi.org/10.15359/ru.36-1.16>
- Andrade, M. (2020). Beyond quality metrics: Defying journal rankings as the philosopher's stone of mathematics education research. *Educational Studies in Mathematics*, 103, 359-374. <https://doi.org/10.1007/s10649-020-09932-9>
- Bastidas, R. (2009). Metodología de la Investigación Educativa (2 ed.). Madrid: La Muralla S.A.
- Bautista, T. (2024). Perspectivas y oportunidades para la integración de analíticas de aprendizaje: un estudio cualitativo en universidades mexicanas. *Revista de análisis de aprendizaje*, 11(1), 49-66. <https://doi.org/10.18608/jla.2024.8125>
- Campos, R. (2022). Analítica del aprendizaje: un desafío al desempeño del personal docente. *Revista Universidad y Sociedad*, 14(6), 40-48. [http://scielo.sld.cu/scielo.php?pid=S2218-36202022000600040&script=sci\\_arttext](http://scielo.sld.cu/scielo.php?pid=S2218-36202022000600040&script=sci_arttext)
- Cuadrado, G. (2020). Gestión del conocimiento en la universidad: cuestionario para la evaluación institucional. *Revista Iberoamericana De Educación Superior*, 11(30), 201-218. <https://doi.org/10.22201/iisue.20072872e.2020.30.596>
- Cuadras, C. (2019). Nuevos Métodos de Análisis Multivariante. Barcelona-España: CMC. <https://es.scribd.com/document/420846198/Cuadras-Nuevos-me-todos-de-estadistica-multivariante>
- Devia, C. (2021). Sistema de gestión del conocimiento basado en estándares de alta calidad para un programa educación superior. *Revista Ingeniería, Matemáticas y Ciencias de la Información*, 8(16), 55-64. <http://dx.doi.org/10.21017/rimci.2021.v8.n16.a102>
- Devine, P. (2022). Diversity training goals, limitations, and promise: A review of the multidisciplinary literature. *Annual Review of Psychology*, 73(1), 403-429. <https://doi.org/10.1146/annurev-psych-060221-122215>
- Díaz, J. (2024). Gestión del conocimiento dirigido a docentes de educación universitaria, desde la perspectiva humanista y transdisciplinaria. *Aula Virtual*, 5(11), 13-28. [https://ve.scielo.org/scielo.php?script=sci\\_arttext&pid=S2665-03982024000100013](https://ve.scielo.org/scielo.php?script=sci_arttext&pid=S2665-03982024000100013)
- Domínguez, D., Reich, J., & Ruipérez, J. (2020). Analítica del aprendizaje y educación basada en datos: Un campo en expansión. *RIED. Revista Iberoamericana de Educación a Distancia*, 23, 33-43. <https://doi.org/10.5944/ried.23.2.27105>
- Escorcía, J. (2020). Gestión del conocimiento en Instituciones de Educación Superior: Caracterización desde una reflexión teórica/ Knowledge management in Higher Education Institutions: Characterization from a theoretical reflection. *Revista De Ciencias Sociales*, 26(3), 83-91. <https://doi.org/10.31876/rcs.v26i3.33235>

- Figaredo, D. D.-V. (2020). Analítica del aprendizaje y educación basada en datos: Un campo en expansión. *RIED-Revista Iberoamericana de Educación a Distancia*, 23(2), 33-43. <https://doi.org/10.5944/ried.23.2.27105>
- Gareth, D., & Trevor, R. (2021). *An Introduction to Statistical Learning: With applications in R*. Estados Unidos: Springer. <https://www.casact.org/sites/default/files/2022-12/James-G.-et-al.-2nd-edition-Springer-2021.pdf>
- Hair, J., Black, W., Babin, B., & Anderson, R. (2019). *Multivariate Data Analysis*. Upper Saddle River: Pearson. <https://www.drnishikantjha.com/papersCollection/Multivariate%20Data%20Analysis.pdf>
- Hernández, R., & Mendoza, C. (2018). Metodología de la investigación. Las rutas cuantitativa, cualitativa y mixta. México: Mc Graw Hill Education. Obtenido de <http://www.biblioteca.cij.gob.mx/Archivos/Materiales de consulta/Drogas de Abuso/Articulos/SampieriLasRutas.pdf>
- Huinga, H., & Méndez, J. (2022). *Gestión del conocimiento, herramientas tecnológicas en la comunicación efectiva en directivos de las instituciones educativas*, Lima-2022. Universidad César Vallejo. <https://hdl.handle.net/20.500.12692/94517>
- Legusov, O. (2022). Equity, diversity, and inclusion from the perspective of international community college students. *Canadian Journal of Higher Education*, 52(4), 121-134. <https://doi.org/10.47678/cjhe.v52i4.189821>
- Martínez, M. (2021). Análisis factorial confirmatorio: un modelo de gestión del conocimiento en la universidad pública. *RIDE. Revista Iberoamericana para la Investigación y el Desarrollo Educativo*, 12(23), 1-23. <https://doi.org/10.23913/ride.v12i23.1103>
- Mendoza, C., & Bullón, C. (2022). Gestión del conocimiento en instituciones de educación superior: una revisión sistemática. *Revista de Investigación en Ciencias de la Educación, Horizontes*, 6(26), 1992-2003. <https://repositorio.cidecuador.org/handle/123456789/2476>
- Mora R. (2015). Factores que intervienen en el rendimiento académico universitario: Un estudio de caso. *Universidad del Zulia*, 31, 1041-1063. <https://www.redalyc.org/pdf/310/31045571059.pdf>
- Mora, C. (2020). En sintonía con esta postura, lo indicado por Mora Yate, Mahecha Escobar y Conejo Carrasco (2020) sostiene que los estudiantes que desarrollan habilidades de autorregulación y cuentan con apoyo institucional tienden a lograr aprendizajes más profundos. *Cultura, Educación y Sociedad*, 11(2), 191-206. <http://dx.doi.org/10.17981/cultedusoc.11.2.2020.12>
- Ocaña, J., Álvarez, I., & Paredes, A. (2023). La gestión del conocimiento e innovación en los procesos educativos en la disciplina del diseño. *Cuadernos del Centro de Estudios de Diseño y Comunicación*, 180, 173-182. <https://doi.org/10.18682/cdc.vi181.8904>
- Rencher, A., & Christensen, W. (2002). *Methods of multivariate analysis (3rd ed ed.)*. United States of America: John Wiley & Sons, Inc. [https://eprints.uad.ac.id/135/1/Handbook\\_of\\_analysis\\_multivariat.pdf](https://eprints.uad.ac.id/135/1/Handbook_of_analysis_multivariat.pdf)
- Ruipérez, J. (2020). El Proceso de Implementación de Analíticas de Aprendizaje. *RIED. Revista Iberoamericana de Educación a Distancia*, 23(2), 85-101. <http://dx.doi.org/10.5944/ried.23.2.26283>
- Sánchez, M., & Martínez, A. (2020). Evaluación del y para el aprendizaje: instrumentos y estrategias. Imagia Comunicación. Ciudad de México: Universidad Nacional Autónoma de México. [https://www.puees.unam.mx/sapa/dwnf/114/3.Sanchez-Mendiola\\_2020\\_EvaluacionDelAprendizaje.pdf](https://www.puees.unam.mx/sapa/dwnf/114/3.Sanchez-Mendiola_2020_EvaluacionDelAprendizaje.pdf)
- Suazo, I. (2023). Knowledge management models in higher education. *European Journal of Education and Psychology*, 6 N° 2, 1-23. <https://orcid.org/0000-0002-1689-1632>
- Suazo, I. (2023). Modelo de Gestión del Conocimiento. *Universidad Autónoma de Chile*, 3(1), 3-14.

<https://repositorio.uautonoma.cl/handle/20.500.12728/10275>

Tabachnick, B., & Fidell, L. (2013). *Uso de estadísticas multivariadas* (6 ed.). New Jersey, Estados Unidos de América.: Pearson Educación.

Tavares, V. (2021). Feeling excluded: international students experience equity, diversity and inclusion. *International Journal of Inclusive Education*, 28(8), 1551–1568. <https://doi.org/10.1080/13603116.2021.2008536>

Van der Velden, G. (2023). Peer-mentorship and first-year inclusion: Building belonging in higher education. *BMC Medical Education*, 23, 833. <https://doi.org/10.1186/s12909-023-04805-0>



Esta obra está bajo una licencia de **Creative Commons Reconocimiento-No Comercial 4.0 Internacional**. Copyright © Dayana Cristina Villarreal Meza y Ciro Diego Radicelli García.

